Models of Upland Vegetation Communities at Multiple Spatial Scales

Liam Butler

Thesis submitted for the degree of Doctor of Philosophy at Newcastle University.

School of Natural and Environmental Sciences, October 2019.

For you,

Mum, Dad and Dianne,

Chris and Darwin

To you,

Roy

General Abstract

The composition and occurrence of vegetation communities changes across multiple spatial scales in response to both environment and human management. Key drivers at small spatial scales (<1m or quadrat-scale) include patch structure between individual species, at intermediate scales (1ha or field-scale) local environmental conditions, whereas at large scales (km or national-scale) broad climate and soil characteristics. This research takes advantage of vegetation data collected via contrasting methods across these multiple spatial scales to quantify the role of these drivers.

Data from 167 1m^2 quadrats in an upland 96ha sheep-grazed heft at Ashtrees Dipper, Northumberland, was used to understand the relationship between vegetation patch patterns and environmental drivers at sub-quadrat (10cm) and quadrat (1m) scales. The numbers, areas and shapes of vegetation patches were primarily determined by soil characteristics, especially pH and water content, and proximity of sheep tracks (distance and length of sheep tracks). The resulting species patch patterns were then interpolated to field scale across the whole 96ha grazing area.

Many countries have developed formal systems to classify vegetation communities, but no single generalisable method exists to allocate vegetation quadrats to community classes. Using the National Vegetation Classification (NVC) as an example, a novel generalisable method was developed to allocate vegetation quadrats to any classification via the computational generation of sets of "pseudoquadrats" for each NVC community at Ashtrees Dipper. These pseudoquadrats were summarised via detrended correspondence analysis (DCA) and new field quadrats placed within the ordination as passive samples. This then allowed a probability score to be calculated for each of the 167 field quadrats for its NVC community membership, which could then be interpolated across the whole 96ha grazing area.

The NVC provides detailed information on the national distribution and characteristics of vegetation in Great Britain. Species distribution models (SDMs) were derived from data in the NVC handbooks, and geographic information system (GIS) predictor layers were used as SDM inputs. Predictions of NVC communities occurring in the protected Biodiversity Action Plan (BAP) habitats in England and Wales were made at 1km spatial resolution. Five SDM

models were tested: generalised linear models (GLM), support vector machines (SVM), random forests (RF), maximum entropy (MaxEnt) and maximum likelihood (MaxLike). The distribution of individual species at 1km scale was then derived from the NVC community predictions. These species predictions were compared to records of species recorded in the National Biodiversity Network Atlas (NBN Atlas), using the catchment of the River Rede, Northumberland (~40km^2) as a case study. GLMs, RF and MaxEnt produced robust predictions of the species distributions, with RF the most accurate.

Overall, this research has demonstrated that the role of environment and management on individual plant species and their communities is best understood at multiple spatial scales, from the influence of sheep grazing in small-scale vegetation patches through to large-scale spatial distributions of species in BAP habitats.

Declaration of Authorship

I declare that the work herein is my own intellectual property. My supervisors, Dr. Roy A. Sanderson and Prof. Stephen P. Rushton, made contributions to each chapter. Any additional contributions or adaptations have been acknowledged appropriately.

Vegetation percentage cover, dominant and subdominant quadrat surveys were originally performed by Dr. Roy A. Sanderson and Prof. Stephen P. Rushton at the Ashtrees Dipper, Northumberland. The data was provided by my supervisors in paper format and were digitised as part of this research. Vegetation percentage cover data, dominant vegetation data and subdominant vegetation was used in Chapter 2. Vegetation percentage cover data was used in Chapter 3. Chapter 4 used vegetation community distribution data from the British NVC handbooks and database. Chapter 4 also used vegetation species presence data adopted from the National Biodiversity Network (NBN) Atlas. The data used has been referenced and acknowledged appropriately.

The research work disclosed in this research project was fully funded by the ENDEAVOUR Scholarship Scheme – Group B – National Funds.

Acknowledgements

Firstly, thanks go to the ENDEAVOUR Scholarship Scheme Group B National Funds, for fully funding this project.

Sincere thanks go to Dr. Roy A. Sanderson and Prof. Stephen P. Rushton for supplying the data as well as offering guidance and support throughout the PhD. Special thanks goes to Dr. Roy A. Sanderson for his continuous leadership, support and for taking me under his wing to make this project a novel approach to vegetation community modelling.

Thanks also go to Dr. Mark Shirley and Dr. Marion Pfeifer for their help and guidance in the research and providing insights into different ways of thinking. Thank you for reminding me that it is important to always be confident in my work, and myself.

To my parents and Dianne, thank you for all your support, emotionally and financially, to pursue a once-in-a-lifetime opportunity. Thank you for always reminding me to never give up, and to think about why I started this journey. Because of you, I have become a strong-willed and better person. I love you.
To you, Chris and Darwin for being there to provide me with patience, integrity and support I needed to keep on going when things got difficult. Thank you for everything.

I would also like to thank my colleagues and friends, especially Sean Gettings, Amy Campbell, Bedour Alsayegh and Samantha De Silva, for the continuous support and help you offered throughout the last 3 years.

Once again, to you, Roy. Thank you for your care and support, and for always being there to hear me out when I was struggling. Thank you for believing in me. Without you, this journey would not have been the same.

To all of you,

Thank you for guiding me to become a man of value rather than a man of success.

To all those who are struggling, remember that

A river cuts through rock, not because of its power, but because of its persistence.

- James N. Watkins

Table of Contents

List of Figures

List of Tables

Chapter 1. General Introduction

1.1 Upland habitats

Upland environments in the UK are formally defined as any environment over 250m altitude and classified by JNCC as 'Natural Areas' that fall within areas where agricultural production conditions are difficult mostly because of poor soil, climate and topography (Backshall *et al.*, 2001; JNCC, 2016). Note that there is a transition between lowland and upland vegetation communities, such that some lowland vegetation may occur in areas formally defined as upland environments (Backshall *et al.*, 2001). In the UK, upland environments are amongst the major contributors to ecosystem services, e.g. drinking water supplies and carbon stores in peat or blanket bogs (Bonn et al., 2009; RSPB, 2015). Sustainable management of these ecosystems and their services is important for both people and wildlife (Kremen, 2005) and as a result, many of these areas are of environmental priority and conservation value.

Amongst the different upland habitats, the research described in this thesis focuses on semi-natural upland habitats dominated by heather moorland and grassland, which are generally managed for grazing or red grouse (Sanderson *et al.*, 1995a; Sanderson *et al.*, 1995b; Cherrill and McClean, 1999; Milne *et al.*, 2002; Stewart and Pullin, 2008). Upland plant communities are affected by biological, physiological and chemical components such as, soil type, topography, meteorology, herbivore grazing, and anthropogenic disturbance (Pott, 2011). Sheep and cattle grazing is an important management tool in upland habitats, and when correctly implemented helps to maintain a diverse mosaic of different vegetation communities (Berg *et al.*, 1997; Adamson *et al.*, 2001; Stewart and Pullin, 2008). However, poor upland management may cause declines in vegetation biodiversity, with subsequent negative effects on some bird populations (e.g. black and red grouse, ring ouzel and curlew; RSPB, 2015).

1.1 Vegetation surveying methods and vegetation classifications

Vegetation is formed of mixtures of different plant species at different abundances, and these are typically described as 'assemblages' or 'communities'. To ensure consistent descriptions of communities amongst scientists many countries have developed standardised national classification systems (Chytrý and Otýpková, 2003; Malik and Husain, 2006). These classifications usually provide a list of plant species found in each class, with their frequency and an indication of cover, and often the associated environmental conditions. The UK has developed a National Vegetation Classification (NVC) as a phytosociological classification system of semi-natural environments in Britain and the NVC provides an extensive floristic dataset of species and communities. Smith *et al.* (1992), Rodwell (1998a), Rodwell (1998b),

Rodwell (2006) and Morecroft *et al.* (2009) describe vegetation communities in the NVC as a holistic classification that adopted a phytosociological approach in developing floristic data sets. In addition to comprehensive information on the species composition of each vegetation community, the NVC handbooks also contain summary information on resource requirements, traits, functional types, species richness and maximum abundance of species. Some countries also have specialised software to classify vegetation sampling data (relevé data) into communities, such as the Modular Analysis of Vegetation Information System (MAVIS) in the UK. Vegetation data has traditionally been collected through random quadrat sampling. Different sized quadrats were used depending on the habitat and vegetation type, for example, $4m^2$ quadrats (2m x 2m) randomly placed in a visually homogenous block of vegetation for habitats with short herbaceous plants or shrubs, $16m^2$ for taller vegetation, sub-shrub heaths and low woodland vegetation, $100m^2$ for tall vegetation, woodland fields and dense scrub or $2500m^2$ for sparse vegetation and woodlands (Rodwell, 1998a). This survey method requires expert knowledge on plant identification, is a very time-consuming task, and thus expensive. In the UK, the NVC is available in extensive handbooks, first published in 1991 (Rodwell, 1998a). These handbooks also provide 10 km x 10 km dot-distribution maps for different vegetation communities but these maps as only cover approximately 80% of the British mainland (Rodwell, 2006). Recently, the NVC classification has been made publicly available in digital format (JNCC, 2016), i.e. the frequency (constancy) and abundance of the species within each community, plus the $10km^2$ resolution maps of national distribution, making the data more convenient for quantitative analysis and prediction.

1.2 Phytosociology

Phytosociology is the scientific study of the composition, phylogeny and relationships between the constituent species that form a plant community (Pott, 2011). In the early Twentieth Century there were two broad philosophical ‘schools’ to describe vegetation communities. The European school, particularly the ideas of Clements (1907) considered the vegetation as forming discrete spatial entities (communities or releves) often changing over time through a successional process (e.g. from pioneer through to climax community). In contrast, Gleasonian theory (Gleason, 1917) suggested that the species composition of a stand of vegetationreflected the collective individual responses of each species to the local environment. . Subsequent research by for example Whittaker (1965) also emphasised this ‘continuum’ interpretation of vegetation, as individual species change. More recently it has been recognised that a synthesis of the two approaches is most valuable to practicing

ecologists, particularly once spatial and temporal scales are considered (Foster et al., 1990). Plant communities tend not to have discrete physical boundaries, unless there is an abrupt change in the underlying environment, but it is nevertheless essential to be able to describe and quantify communities.

Environmental change, e.g. due to climate change, land management etc., can alter relationships between neighbouring species, and hence affect vegetation community class (Van der Maarel, 1979). The distribution of plant species and the communities they create depend on favourable environmental conditions that align with the requirements for those species (Pott, 2011). Plants compete with each other and form different patches depending on the environmental resources available (Crain and Bertness, 2006). Indeed landscapes are generally described in terms of patches, vegetation corridors or species matrices (mosaics; Cushman et al., 2010). Vegetation patterns are characterised by size, shape and spatial distribution, whilst different types of patterns arise from factors such as soil pH, soil type, drainage, slope and grazing (Austin, 2002; Stewart and Pullin, 2006). The spatial scale of observation may affect what patterns are visible, for example within a quadrat, at quadrat or field scale, and through to landscape or national scale. The latter in particular may relate to broad environmental conditions (Chytrý and Otýpková, 2003). This can provide a basis for upscaling vegetation-environment relationships from small to large spatial scales, especially with the availability remote-sensed data (Zak and Cabido, 2002; Chytrý and Otýpková, 2003). Landscape modelling of plant communities could theoretically provide insights into community distributions at different spatial scales if the most important environmental drivers are considered (Franklin, 2010). Unfortunately, it can be harder to quantify the relationships between environment and rare vegetation communities with restricted national distributions. This can make it more difficult to design conservation strategies to protect them from anthropogenic threats (White, 1979). Policy-based strategies to protect ‘rare’ vegetation communities would require a solid base on the communities’ known occurrences, probability predicted presences as well as deeper validation of the outputted models.

1.3 Ordination techniques

This research, especially Chapters 2 and 3, uses different ordination techniques to analyse and visualise species data. Ordination analysis are multivariate techniques that evaluate and search for patterns across species composition within sample data (Zelený, 2017). These multivariate techniques summarise complex responses by identifying gradients between species

composition and uses these gradients to explain the relationship with environmental variables. In addition, ordination techniques reduce such complex responses into two (or three) axes that best fits the data (Zelený, 2017). The main ordination methods used in this thesis were Detrended Correspondence Analysis (DCA), Non-metric multi-dimensional scaling (NMDS) and Redundancy analysis (RDA). In general, the selection of the most appropriate ordination technique depends on the whether environmental data is included in the analysis and the characteristics of the vegetation data (Lepš and Šmilauer, 2003). Where an ordination is undertaken without environmental data (or the environmental is merely used to aid interpretation afterwards) it is referred to as an 'unconstrained' method. In contrast, where the environmental data are used simultaneously within the ordination, so as to change the resulting ordination scores, it is referred to as a 'constrained' method (Jongman, 1995).

Where vegetation is characterised by a large number of ubiquitous species, and a smaller number of infrequent species, so-called 'linear' ordination methods are often appropriate. In these methods, such as principal components analysis (PCA - unconstrained) and redundancy analysis (RDA – constrained), species abundances are assumed to increase or decrease monotonically along each ordination axis, in an approximately linear fashion. As the number of ubiquitous species declines, and the number of infrequent species species increases, the data becomes more 'sparse', such that the species by samples matrix becomes dominated by zero values. In this situation so-called 'unimodal' methods are sometimes better to characterise the vegetation, for example correspondence analysis (CA – unconstrained) and canonical correspondence analysis (CCA – constrained). On each ordination axis the abundance of a species can rise and fall, as appropriate, such that more flexible, Gaussian (bell-shaped) curves are fitted rather than straight lines (Legendre and Legendre, 2012).

One problem that can arise in unconstrained ordinations, especially CA, is an 'arch-effect' or 'horseshoe effect' in the resultant output, where the ends of the sample points on the first ordination axis start to occur close together in ordination space, even though their species composition is relatively dissimilar. This artefact arises when the species turnover in the dataset is relatively high (Legendre and Legendre, 1998). Detrended Correspondence Analysis (DCA) is a unimodal unconstrained ordination method that tries to correct this artefact by subdividing the first ordination axis into segments, and rescaling the scores (Hill, 1979a). Hill and Gauch (1980) note that the length of the first DCA ordination axis is expressed in multiples of the standard deviation. This means that a Guassian response curve rises and falls after about 4 s.d. units, i.e. samples that differ by more than 4 units on the first

axis probably have very few species in common. Thus, DCA provides an indication of whether the data set is truly unimodal or, otherwise, linear based on the length of the gradient of the first axis (Zelený, 2017). An alternative unconstrained method is non-metric multi-dimensional scaling (NMDS) which is based on a non-metric measure of dissimilarity between sample points and is generally restricted to low dimensions (2 or 3 axes). The NMDS algorithm uses iterative shuffling of samples to find the best distribution amongst points.

The results of constrained ordination methods such as RDA and CCA can be displayed in the form of a 'biplot' (species and environmental variables) or a 'triplot' (survey sites, species and environmental variables) to diagrammatically represent the relationship between all three factors (Lepš and Šmilauer, 2003; Zelený, 2017). These types of plots include arrows the length of which can infer the importance and effect size of each variable and their directions the correlation between variables. Constrained methods also allow significance testing, via permutation tests in which the samples and environmental data are shuffled, and the randomised data compared with the actual data. These allow (pseudo) F-ratios and p-values to be calculated for the effects of each environmental variable on the community composition (Jongman et al., 1987).

1.4 Modelling, Evidence and Policy

It is important to understand the nature of a complex ecological system by using evidence-based scenarios, embedding knowledge and data into realistic model outputs. The overall modelling and evidence frameworks then feeds into policy, allowing for better decision making and setting of targets. Policies that address habitat conservation include the Strategic goals and Aichi targets in the Convention of Biological Diversity (CBD; United Nations, 1992), the European Habitats Directive (92/43/EEC; European Commission, 1992) and the European Strategic Environment Assessment Directive (2001/42/EC; (European Commission, 2010). The UK Biodiversity Action Plan, updated to the UK Post-2010 Biodiversity Framework (JNCC and DEFRA, 2012) aims to implement policies that address challenges to reduce biodiversity loss, conserve nature, stop environmental degradation and support ecological networks for sustainable growth (JNCC and DEFRA, 2012).

Robust and reliable qualitative and quantitative models can be an important tool to develop policy frameworks. These models aid creation of realistic, achievable set targets to be reached within a specified time period (Bartlett and Kurian, 1999; Boulanger and Bréchet, 2005).

Different models can be applied depending on the policy-framework, conservation efforts, monitoring schemes and can thus be question-specific. In some cases models can be used to predict possible locations of species, both for conservation and if necessary pest control (Lambin et al., 2014). However, it is important to note that ideally such models should be validated, and also provide information on their accuracy and precision. Best practice is to develop environmental policies in tandem with models to address environmental issues and improve the quality of decision-making (Bartlett and Kurian, 1999). A reliable evidence-base, such as expert knowledge, field data and previously published scientific literature, is required to both set these targets for the environment and monitor progress in reaching them. Such evidence-bases can be derived from the scientific literature and expert knowledge to develop study designs and appropriate models (Stewart and Schmid, 2015). One drawback that may be encountered in reaching the set targets is the high financial and human effort required to monitor the environment (Ball, 1999). The availability of high-resolution remote-sensed and environmental data has the potential to reduce the financial costs of measuring ecological health by developing predictive and explanatory models to help identify areas of vegetation or broader landscapes in need of conservation action. Such models allow for identification of areas that are of poor environmental status more effectively and if necessary more rapid implementation of appropriate remedial management. It is also important that scientists are able to communicate research findings to non-experts. For example, within the UK, the Parliamentary Office of Science and Technology (POST) publish a series of 4-page summary documents ("POSTNotes") available from the Government website (www.parliament.uk/postnotes). Whilst there are indications of some potential biases in these summary documents (Seton-Clements et al., in prep), they nevertheless provide a valuable resource for policy-makers.

1.5 Study site

This research makes extensive use of the Ashtrees Dipper grazing 'heft' within the catchment of the River Rede in Northumberland National park, UK (Fig. 1.1). The Ashtrees Dipper is a 96ha field site primarily dominated by rough grassland and moorland habitats (Rushton *et al.*, 1992; Smith *et al.*, 1992; Sanderson *et al.*, 1995a). It is primarily sheep-grazed and contains a number of drainage ditches at the southern side. It is a north-facing site, with altitudes ranging between 250m and 350m and is characterised by acidic soils and overall high water content.

The species data used in this research (in Chapters 2 and 3) was obtained from a field survey performed in 1991 at the Ashtrees Dipper (Roy Sanderson and Stephen Rushton, pers. comm.). In addition to these vegetation surveys, environmental data including soil pH, soil water content and slope were taken *in situ*. 167 1m^2 gridded wire quadrats placed along transects 150m apart, with each quadrat 75m apart along each transect (fig. 1.1). Two types of vegetation surveys were performed, i) conventional visual vegetation percentage abundance of all the species of plants (including Bryophytes) within a whole 1m^2 quadrat and ii) dominant and subdominant vegetation identification within each 10cm quadrat grid cell. The latter was done by visually identifying the most abundant species (dominant) and the second most abundant species (subdominant) within each 10cm grid.

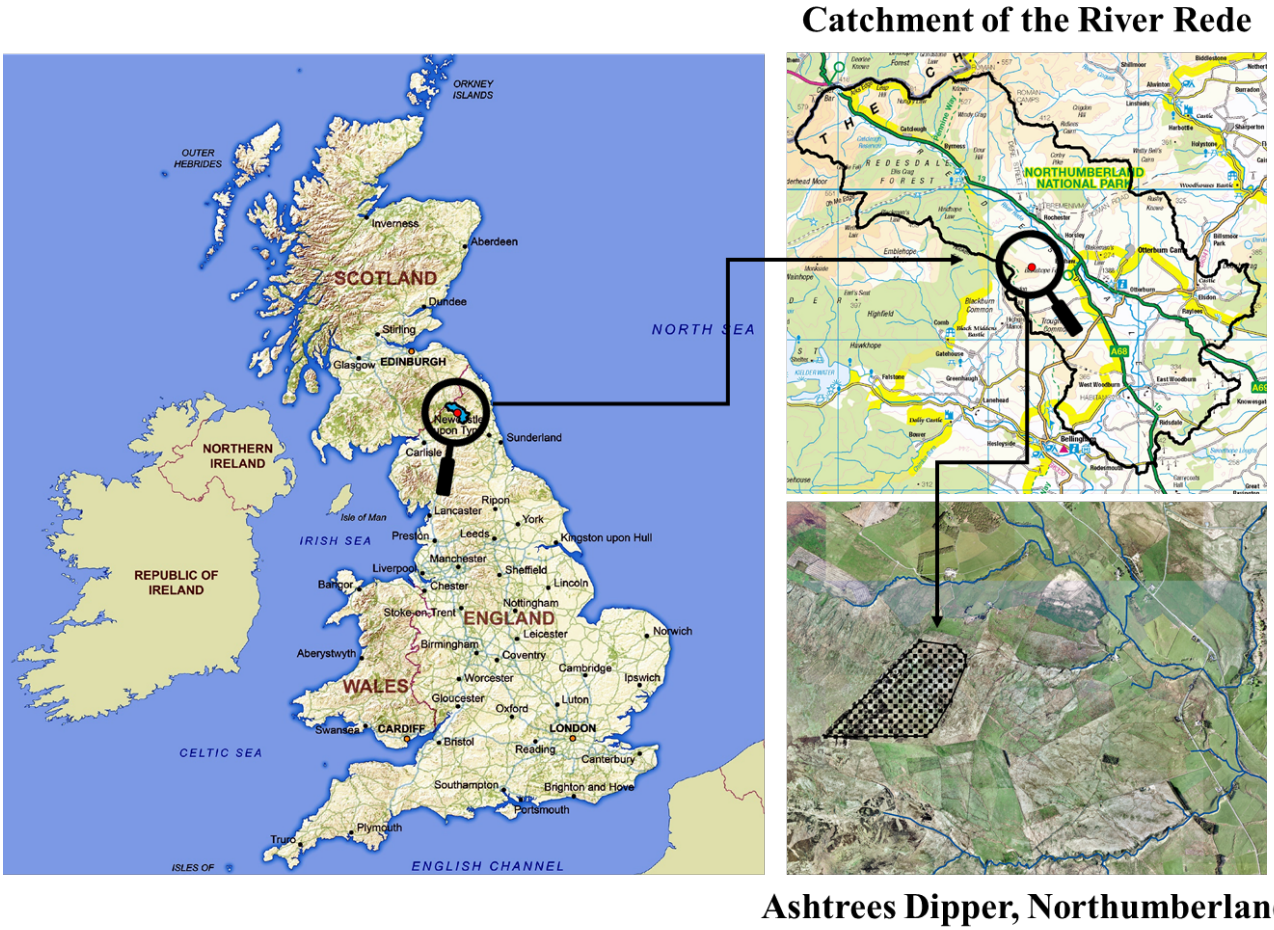


Fig. 1. 1 Location of the study in the NE UK (left; blue polygon = Catchment of River Rede, red point = Ashtrees Dipper), ii) close up of the Catchment of the River Rede at Northumberland National Park (top right; black polygon = Catchment of River Rede) and iii) location of the 167 surveyed quadrats (points) at Ashtrees Dipper, Northumberland National Park, Northumberland, UK (black points = position of quadrats; blue line = water course and River Rede).

1.6 Thesis Rationale and outline

Scientists encounter several challenges in understanding the key factors that affect vegetation at multiple spatial scales. These includes patch structure of individual species, the interaction between species as well as and the relationships with both the biotic and abiotic environment (Parkes *et al.*, 2003; Fischer and Lindenmayer, 2007). This research uses a range of vegetation and environmental data collected at different spatial resolutions to understand and predict how vegetation species and communities interact with each other and the surrounding environment.

Vegetation was assessed at the individual species level from field data collected at sub-quadrat and quadrat scale. This allowed analysis of vegetation patch metrics in relation to the environment, and also facilitated comparison between the sub-quadrat and conventional whole-quadrat survey methods (Chapter 2). Availability of environmental data across the whole of Ashtrees Dipper allowed predictions of both patch metrics and species at the field scale (Chapter 2). To model vegetation community distributions regionally or nationally it is important to have a method to allocate quadrats to vegetation communities (here the NVC classes). While algorithms and software exist to classify groups of quadrats collected using standard NVC protocols, there is no generalisable method to classify single quadrats. This research develops and tests an approach to classify quadrats that have been collected via any typical protocol different to the NVC. This approach can be sufficiently generalisable and used internationally, especially in cases where a vegetation classification system is not currently in place (Chapter 3). Finally, methods were developed to predict upland UKBAP NVC communities across the whole of England and Wales at 1km resolution, and these were compared with observed NBN Atlas records for the associated species in the Rede Catchment (Chapter 4). This thesis explores vegetation-environment relationships across multiple scales, from 10cm to national by understanding the different relationships between species and the different environmental factors involved within each spatial level.

1.5.1 Chapter 2: Understanding vegetation species patches at small spatial scales

The first data chapter uses vegetation data collected at 10cm sub-quadrat scale within each of the 100 cells of a $1m^2$ quadrat identifying the dominant and subdominant species in each cell. Conventional visual percentage cover assessments of all species in the $1m^2$ quadrat were also available. By aggregating the 100 records for the dominant or subdominant species it is

possible to compare the two data collection methods, by NMDS ordination of each dataset, and comparison with Procrustes rotation. Patch metrics for dominant and subdominant vegetation were calculated and these used to create maps across the whole field. Therefore, understanding vegetation formation at sub-quadrat scale provides additional insights into the responses of vegetation to the environment (Greig-Smith, 1979).

Identifying dominant and subdominant vegetation at the species level can also help scientists understand and evaluate vegetation patch fragmentation that is less obvious when surveyed at larger scales (Greig-Smith, 1979). Understanding the process and drivers of patch fragmentation at very small scales (e.g. 1m) provides additional scientific insights into how, for example, vegetation species influence each other and grow together. Furthermore, such patchiness can be driven by above- or below-ground processes that have a large underlying effect. This research, however, does not delve into below-ground processes in terms of patchiness of vegetation species but rather looks at species in, effectively, 2-D space and tries to identify possible interactions based on their patch-forming capabilities.

1.5.2 Chapter 3: Developing generalisable methods for vegetation classification using computational 'pseudo-quadrats'

The second data chapter in this research aims to develop a novel method to classify individual vegetation quadrats to communities in the NVC. Whilst software such as the Modular Vegetation Analysis Information System (MAVIS; Smart *et al.*, 2016) is already available, it assumes that quadrats have been surveyed using standard protocols. In the case of the NVC this is 5 randomly placed quadrats ($4m^2$) within a visually homogenous block of vegetation. However, there are large amounts of extant vegetation records collected using other methods, typically $1m^2$ quadrats along transects or similar. Two complementary approaches are described in this chapter to create artificial 'pseudoquadrats' of NVC communities that can be analysed through conventional multivariate ordination methods. Field quadrats, collected by any suitable method, can then be placed within this ordination space, and their NVC class determined. Probabilities of communities for each quadrat in a transect can be interpolated, to create predicted maps for each community at the field scale. The use of ordination techniques allows for better interpretability, flexibility and efficiency when compared to conventional models (Van der Maarel, 1969; Hui *et al.*, 2015). The developed methods of using ordination methods

with computer-generated pseudoquadrats helps fill in data gaps as well as provide a reliable assessment of possible vegetation communities, even at very low spatial scale.

1.5.3 Chapter 4: Predicting and assessing vegetation community distribution models across England and Wales

The third and final research data chapter aimed to predict vegetation communities at a very large (national) scale. This chapter used environmental and remote-sensed data at high resolution to predict vegetation community distributions using species distribution models (SDMs, Franklin, 1995; Franklin, 2010) at national scale for the upland UKBAP communities in England and Wales. Vegetation communities are difficult to assess and identify at very large spatial scales, mostly because of the costs and human effort. Using SDMs prediction of vegetation communities can be a major benefit for scientists and policy makers to assess vegetation community health in areas high environmental concern and protection **(**UKBAP; JNCC, 2016). Vegetation abundance is independently being collected as part of a national plant recording scheme across the UK (Pescott *et al.*, 2015; Pescott *et al.*, 2019). However, identification of suitable sites on which to focus monitoring of vegetation species is difficult. Field monitoring can provide important baseline data on which to base policy decisions, for example collected every 10-12 years via detailed surveys of 1 km samples in the UK Countryside Survey. However, detailed field monitoring is often expensive, and if an aim of the monitoring is conservation of specific communities or taxa is then expert knowledge on locations where these are likely to occur or are threatened is needed. Prediction of probable locations of vegetation communities can help address this issue by using modelling frameworks, such as presented in this thesis, e.g. via species distribution models (SDMs). It should be noted that such models, while evaluated for accuracy, still have some difficulties in validation (see Chapter 4).

Information on vegetation community distribution in this research can provide a link to help address issues in locating vegetation communities, and can be extended to identify species occurrence probabilities, potentially aiding identification of monitoring sites. This approach is tested in Chapter 4 using NVC predictions across England and Wales, with special focus at the regional scale (the catchment of the River Rede in Northumberland, UK) using data from the NBN Atlas for semi-qualitative evaluation.

1.7 Overall thesis aims and objectives

The overarching aim of the project is to assess vegetation patch patterns at very small spatial scales (sub-quadrat and quadrat scale), develop generalisable methods to classify vegetation into communities in cases where abundance data does not conform to the conventional survey method and predict spatial distribution of upland vegetation communities at regional and national scale. Fig. 1.2 provides a schematic summary of each data chapter and overviews the methods used for modelling vegetation at different spatial scales.

Specific objectives include:

Chapter 2

- compare traditional survey methods with surveys of dominant/subdominant vegetation
- quantify vegetation patterns formation at sub-quadrat (10cm), quadrat (1m) and grazing heft (1ha)

Chapter 3

- develop robust methods of allocating field quadrats within the standard framework of the National Vegetation Classification

Chapter 4

- Predict UK BAP upland vegetation community distributions using species distribution models (SDMs)
- Use citizen science data from the NBN Atlas to assess reliability of the predicted models

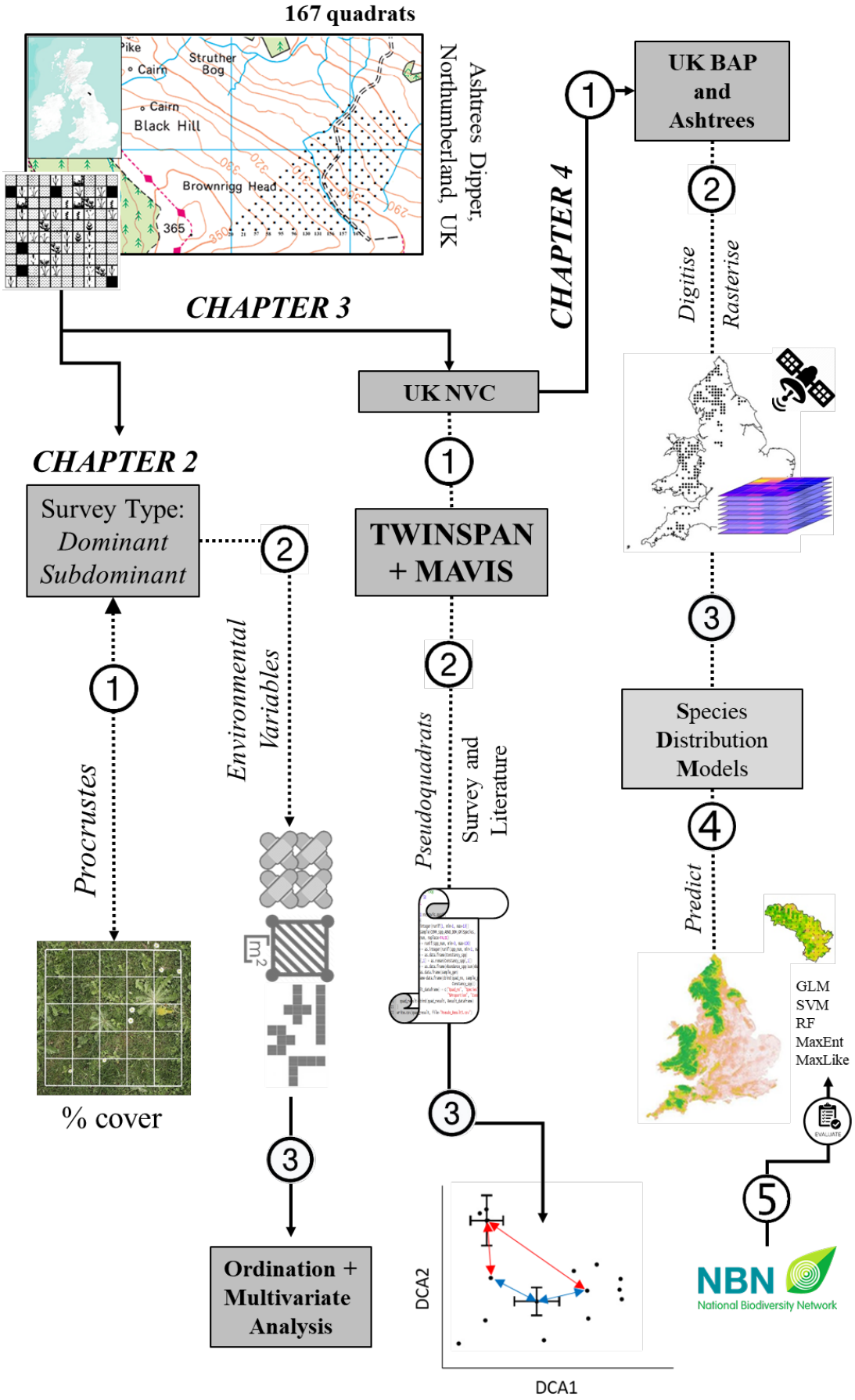


Fig. 1. 2 Graphical overview of the thesis structure and data chapters.

1.8 The link between chapters: a multiscale approach (Fig. 1.3)

Within a community species independently and collectively require different environmental resources. Measurement of vegetation patch structure at small spatial scales provides data that can be used to determine how species respond at the level of clumps (1m) and potentially individual plants (10 cm) to the environment, model growth patterns and interactions between species, as well as link to functional traits. While this part of the research is site-specific (Ashtrees study site, Northumberland), it can still provide insights into the broader patch-patterning of different species in relation to environmental resources. The inclusion of spatial distribution of 'sheep tracks' as a surrogate for grazing pressure on vegetation patches, and positions of drainage ditches, provides additional insights into both grazing and management practices. This is a relatively novel approach at the small spatial scales used to quantify vegetation at quadrat or subquadrat scale when analysing grazing across a whole field. Moving from a small spatial scale (10cm or 1m) would then require identifying clusters of different vegetation species as an assemblage to fit within the NVC-level scale of vegetation classification.

The UK and many other nations have developed systems, often using custom-written computer software, to allocate vegetation samples into their national description of vegetation assemblages. This thesis describes a much more generalisable method to classify vegetation samples, typically from quadrats, which is flexible enough to be used in any current national classification system. This generalisable system uses ordination techniques which offers a two main benefits: firstly ordination provides a good visualisation of the overall patterns amongst the different vegetation communities and secondly the distance between each community in ordination space provides an indication of how similar communities are in their species composition. New data, from for example field quadrats or similar sampling technique, can be placed within this ordination framework of plant communities, and the probability of these quadrats belonging to each plant community can thus be calculated. Furthermore, the spatial locations of the original field survey data used to create the National Vegetation Classification is reported in the NVC Handbooks, and is now also available electronically. This permits the distribution of the communities to be predicted using 'species' distribution models (SDM): this thesis demonstrates that SDM can be successfully modified to predict the districtuion of communities rather than species. SDMs require reliable environmental data, such as elevation, soil type etc., to make the predictions, and the original data must be divided into test and training datasets, to avoid spurious inferences of the true model accuracy.

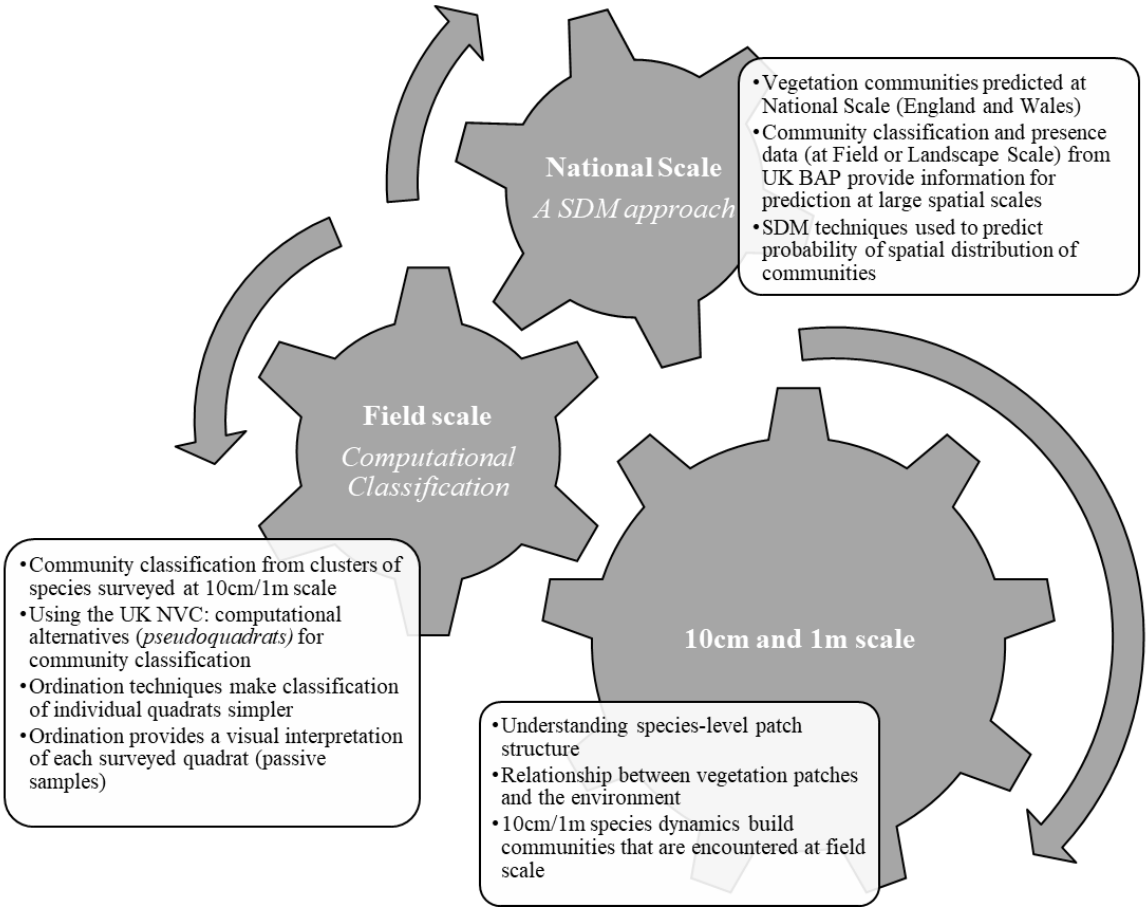


Fig. 1. 3 The clog system of this research a) *Clockwise rotation* = bottom-up approach (From Species level to Field scale to National Scale) and b) *Anticlockwise rotation* = top-down approach (from National scale to Field Scale to Species level)

1.9 Broader scope of the research

The research presented in each chapter in this thesis holds potential to support and add to conservation efforts in both research and evidence-based policy-making. The research provides a coherent narrative through the chapters to inform scientists of key processes affecting plants from the scale of 10 cm through to landscapes. At the same time, each chapter can be utilised independently of the others by different end-users, depending on their requirements. For example, the first data chapter looking into vegetation at sub-quadrat/quadrat scale allows scientists to use the 'dominant/subdominant' survey technique rather than the traditional visual abundance estimation technique to quantify the overall species composition within an area. The patch structure analyses that are made possible from 'dominant/subdominant' surveys, are easy to generalise to any comparable semi-natural vegetation at other sites. If adopted more broadly as a technique, it has the potential to provide

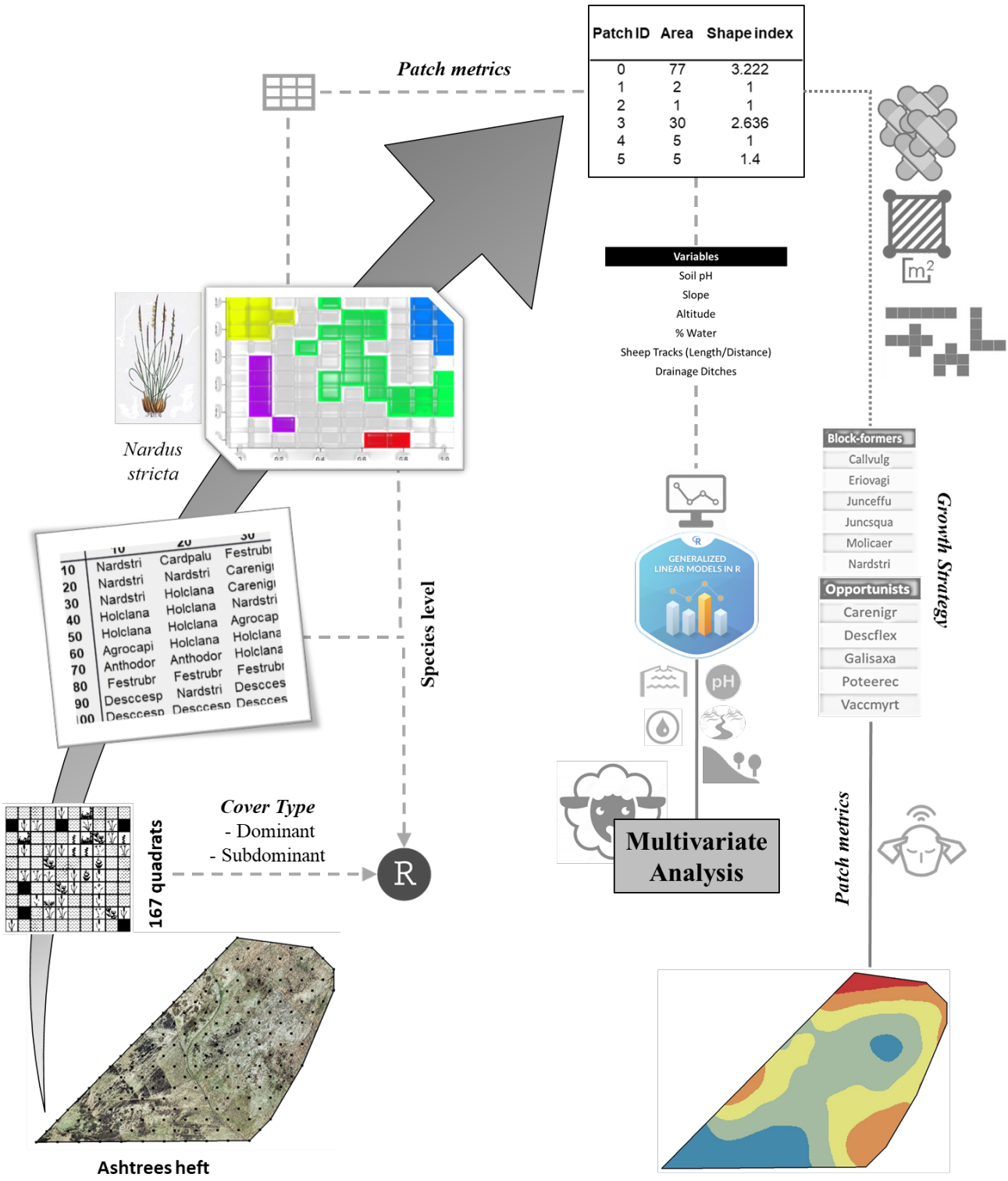
a robust method of detecting differences in patch structure even between sites that are relatively similar in species composition.

The second data chapter describes a generalisable method for vegetation classification that can be used by nations that have no extant classification system in place. The research is also of value in situations where publically available software to classify field quadrats to communities has yet to be developed. The use of ordination techniques, pseudoquadrats and survey data as passive samples also allows the end-user to allocate newly surveyed quadrats to their vegetation classification. Where the pool of potential communities in which field survey vegetation is likely to be limited, the pseudoquadrat method, combined with ordination, provides both a ready means of restricting potential predictions to those communities that are most plausible, and aids interpretation by visualisation in a conventional ordination framework.

The final data chapter provides insights into the use of these methods for policy-makers and conservation research. While the data used in this chapter is at a relatively coarse spatial resolution compared to the preceeding chapters, the results nevertheless provide a solid basis on probable occurrences of protected vegetation communities. There is also scope for different methods to predict vegetation communities using community constituent data (species data) rather than purely community occurrence data. In either case, the community predictions can still be used as a baseline for policy-makers as indication of the possible locations of UK BAP vegetation communities.

Chapter 2. Understanding vegetation species patches at small spatial scales

Graphical Abstract



Abstract

Vegetation patterns are characterised by size, shape and spatial distribution of the patches created by the growth of individual plants within a community. In this research, vegetation patches are defined as a matrix or mosaic of species of plants that can form clumps or individual strands. In upland semi-natural vegetation at the field scale, different types of vegetation patterns have been attributed to soil hydrological and grazing patterns by livestock. Vegetation is typically measured at larger scales than individual plants or clumps of plants, for example, a standard $1m^2$ quadrat is likely to include numerous patches and individuals. Surveys at this spatial scale do not permit a deeper understanding of the ecological processes that might affect individual plants or clumps, nor how metrics derived within a quadrat scale-up to a whole quadrat. The aim of this chapter is to understand the spatial distribution of patches formed by dominant and subdominant vegetation species at very small scale (10cm), the environmental factors that affect the vegetation patches (summarised via patch metrics), and the patch-environment relationships when scaled-up to the whole $1m^2$ quadrat.

Patch metrics analysed were i) total patch number per species per quadrat, ii) patch area per species per quadrat and iii) shape index per patch per species per quadrat. Dominant and subdominant species data, surveyed at a 10cm subquadrat-scale within 167 $1m^2$ quadrats, was used to analyse patch dynamics across Ashtrees Dipper, Northumberland. Results indicated that vegetation patterns are affected by the environment, especially soil pH, altitude, slope and soil water content. The number of patches formed by dominant and subdominant species are, overall, affected by distance to sheep tracks within a 10m buffer. The area of dominant vegetation patches is affected by distance to sheep tracks within a 35m buffer. The shape of patches formed by dominant species is significantly affected by nearby drainage ditches. The area and shape of patches formed by subdominant vegetation, are not significantly affected by proximity of sheep tracks and distance from drainage ditches. Number of patches for some individual dominant vegetation species, e.g. *Nardus stricta* and *Eriophorum vaginatum* are positively related to slope and soil water content respectively. Patch shape of *E. vaginatum* and *J. effusus* show a significant negative influence by slope and altitude respectively. Number of patches formed by subdominant *Molnia caerulea* and *E. vaginatum* are affected by soil characteristics (pH and water content, respectively) while only patch area formed by *D. flexuosa* is significantly affected by soil water content. Shape of subdominant *N. stricta* shows a significant negative correlation while *E. vaginatum* shows positive correlation. None of the

‘block-forming’ or ‘opportunistic’ vegetation species had significant relationships with distance to and length of sheep track or distance to nearby drainage ditches. Results also indicated a close accord between vegetation measured at the detailed sub-quadrat scale with measures of community composition derived from visual estimates of percentage cover of the whole quadrat. This suggests that the environmental processes that affect the vegetation patterns at very small spatial scales are reflected in coarser statistics collected at larger spatial scales.

2.1 Introduction

Environmental drivers have the potential to affect vegetation at many spatial scales, from that of an individual plant, a clump of plants of the same species, mixed patches and plant communities, through to habitat and landscape-level. Most ecological surveys have used one spatial scale, for example the traditional quadrat for community surveys, or field surveys and remote-sensed data for habitats. However, plants grow and compete with each other at much smaller spatial scales, for example, sub-metre patches comprised of a small number of species. It is important to determine the processes that occur at these small spatial scales especially for deeper phytosociological understanding of the relationship between species and the environment.

Vegetation patterns are characterised by size, shape and spatial distribution of plants forming distinct patches of relatively high or dense cover. These patches can be viewed as occurring in a matrix of other species of plants, which occur at lower densities and do not necessarily form distinct patches. There are few studies at small, within-quadrat scales, with most focussing at larger field or landscape scales; e.g. Aguiar and Sala, (1999) reported shapes of dense patches is variable, and may form distinct bands or sometimes irregular or generally circular clumps in shape at scales over 100m. The different types of patchiness have been attributed to variation in hydrological networks (e.g. drainage) and, in semi-natural landscapes, grazing effort by herbivores (Aguiar and Sala, 1999). The upscaling techniques described in this chapter can provide additional insights into patch patterns of vegetation species commonly found within heather moorland and upland grasslands across NE England. The vegetation in these areas has been substantially modified through livestock grazing and changes to hydrology due to digging of drainage ditches (‘grips’) after the Second World War.

Different landscape metrics have been used across spatial scales, i.e. individual patch-level metrics can potentially be used in lower spatial-scales or compared to higher scales of ecological information (McGarigal, 2017). For example (Aguiar and Sala, 1999) describe and measure patches in term of bands, irregular or circular in overall shape stating that banded vegetation patches tend to maintain their shape, growing laterally, while circular/irregular patches show no unidirectional overall pattern. Other patch classification or assessments have used ‘fractal geometry’ emphasising fractal fragmentation of vegetation patches, clusters, and fractal spatial patterning (Sugihara and May, 1990; Li, 2000; Liu and CHEN, 2000). Fractal

geometry, however, is a complicated measure of patch structure whereby there are an infinite number of dimensions and the object to be measures is continuously jagged within every dimension (Marion, 2015). Since this research looks at patch structure in the simplest sense, fractal dimensions where thought to be too complex and thus a simpler form of measure of patch metrics was used, namely the inclusion of shape index as a substitute for vegetation patch shape measure (Section 2.1.3; McGarigal, 2017).

Measurement of vegetation metrics can assess changes in patch development at different spatial scale (centimetres, metres, hectares, kilometres) and can be driven by resource availability, herbivore grazing and species competition (Ritchie, 2009). Semi-natural vegetation often creates a patchy discontinuous structure, for example as a result of water availability (Aguiar and Sala, 1999) and topology (Klausmeier, 1999; von Hardenberg *et al.*, 2001). Vegetation patterns and patches of individual species are variable, sometimes classified as ‘stripes’ or ‘bands’, ‘spots’ and ‘clumping’ (Aguiar and Sala, 1999; Klausmeier, 1999; von Hardenberg *et al.*, 2001). While these descriptions are given to vegetation patterns at larger spatial scales (landscape), patch metrics using dominant and subdominant vegetation species at the quadrat or sub-quadrat scales can allow understanding of how each species interact with each other and environmental drivers (van der Maarel, 1988; von Hardenberg *et al.*, 2001; Ritchie, 2009). It should be noted that for the purpose of this research, dominant vegetation refers to vegetation species that is most visually abundant within any given cell while subdominant vegetation refers to the second most visually abundant species. As such, dominant and sub-dominant simply refer to cover-abundance, not a specific plant trait.

2.1.1 Vegetation at quadrat and sub-quadrat scale

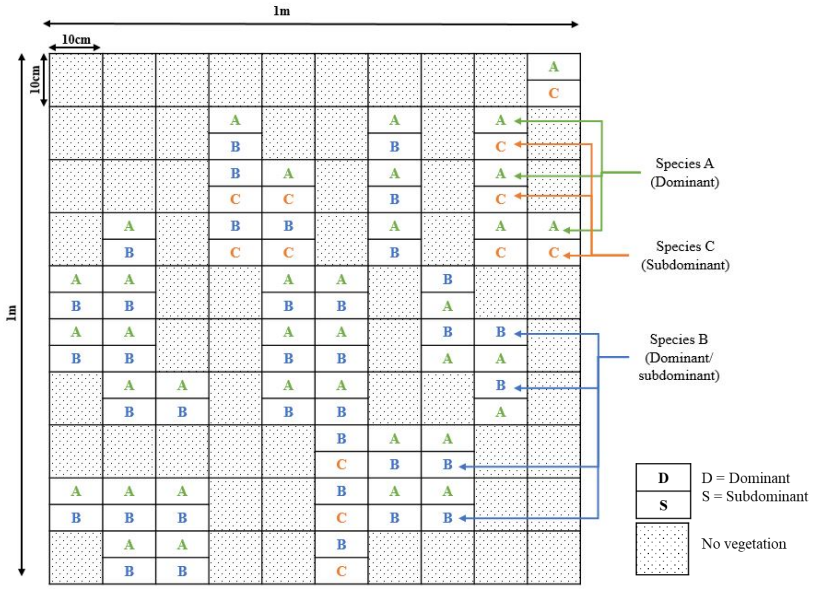


Fig. 2. 1 A visual representation of a $1m^2$ quadrats divided into 10cm cells (sub-quadrats). Each cell hosts a dominant and subdominant species. For ease, dominant vegetation is placed at the top of each 10cm cell and subdominant vegetation is at the bottom of each 10cm cell. E.g. Species A is dominant in 3 cells and Species C is subdominant in those same cells. Species B is dominant in 2 cells and subdominant in another 2 different cells.

Investigating vegetation patterns across a field using $1m^2$ scale quadrats and 10cm cells within each quadrat will provide a better understanding of some selection pressures acting on species as they compete for resources (Ritchie, 2009). Some species may have similar patch metrics in different areas of the field, whereas others may respond to changes in their environments such that their patch metrics differ across a field. If the external environmental drivers that affect the patch metrics of each species and group of species are quantified, then it should be possible to predict the patch metrics anywhere on a field (van der Maarel, 1988; Ritchie, 2009). This chapter uses $1m^2$ quadrats divided into 10cm x 10cm cells, henceforth referred to as a 'sub-quadrat'. In different scenarios, one species, e.g. Species A can be dominant in 3 out of 100 cells while Species C is subdominant in the same 3 cells. Species B, however, is dominant in 2 cells and subdominant in 2 different cells (Fig. 2.1).

2.1.2 Patch metrics

The traditional method for vegetation surveys is visual estimation of percentage cover of each species that occurs within a quadrat of any size (e.g. $1m^2$ as in this research or $4m^2$ for NVC classifications; Rodwell, 1998a). However, for this part of the research the measure of what the dominant and subdominant species are within each cell within a quadrat was used, and the

study compares the reliability of using dominant and subdominant vegetation with traditional percentage cover surveys. Using dominant and subdominant vegetation within each cell also provides better insights into responses of vegetation to environmental influences at a very small scale. Previous literature (Watt, 1947; Greig-Smith, 1979; van der Maarel, 1996; Debinski, 2006) has broadly discussed the importance of assessing vegetation at small-scale without implicit implication of assessing vegetation on a cell-by-cell basis. One limitation of this method is that it does not detect rare species that are neither dominant nor subdominant in any given sub-quadrat but are detected by a percent abundance survey. However, the use of a dominant and subdominant technique might be a quicker way of identifying vegetation communities at any given scale.

Identifying dominant and subdominant vegetation at small scales can aid interpretation of vegetation patch fragmentation that may not be obvious from large-scale surveys (Greig-Smith, 1979). Patch metrics of vegetation at the species-level might translate to patterns in the vegetation community at higher scale. While environmental factors might not affect all vegetation species at quadrat or sub-quadrat scale, major environmental change is likely to alter the dominant and subdominant species. The dominants and sub-dominants are primary species that define a typical vegetation community, which then affects higher spatial levels such as overall landscape habitat (Milchunas and Lauenroth, 1993; Rodwell, 1998a; Rodwell 1998b; Rodwell, 1998c; Rodwell, 2006).

The main environmental drivers at Ashtrees Dipper identified from previous studies are soil pH, soil water content, altitude, slope and sheep grazing (Rushton *et al.*, 1992; Sanderson *et al.*, 1995a; Adamson *et al.*, 2001). It is assumed that dominant and subdominant species form mosaics at any given spatial scale that interact with neighbouring patches to form co-existing species that use similar resources. Three patch metrics of particular value in this context are i) total patch number per species per quadrat, ii) patch area per species per quadrat and iii) shape index per patch per species per quadrat (Addicott *et al.*, 1987; Ritchie, 2009; McGarigal, 2017)

2.1.3 The importance of Shape Index (SI)

The shape index is a form of diversity index to quantify habitat patterns and provides an alternative to fractal dimensions in landscape ecology (Ritchie, 2009; McGarigal, 2017). It measures the complexity of a two-dimensional object compared to a standard shape, generally

a square or circle, of the same size, and therefore alleviates the size dependency problem of the perimeter-area ratio. In its basic form, the shape index of any land parcel or object is a numerical quantity representing the degree to which a shape is compact and it has therefore also been described as a compactness index (Dijkstra *et al.*, 1994; Li *et al.*, 2013). The shape index, *SI*, is:

$$SI = \frac{length\ of\ perimeter\ of\ patch}{length\ of\ perimeter\ of\ a\ square\ of\ same\ area}$$

Thus, the index equals 1 for square patches of any size and increases without limit as the patch becomes increasingly non-square (i.e., more geometrically complex, Fig. 2.2; McGarigal, 2017). Shapes can also be compared to a circle, in which case the formula is derived from the ratio between area of the shape, *A*, and the area of the smallest relatable circle, A_{sc}. However, this method is less widely used due to the problem that shapes with holes or breaks cannot be related to a 'true' circle and becomes scale-irrelevant (Li *et al.*, 2013). Another problem using circles rather than squares as the reference shape is that if the shape of the object is either very compact or distortedly elongated values <1 can be obtained, making interpretation more difficult. The reference shape used in this thesis is a square, hence the minimum possible value of any patch is 1.

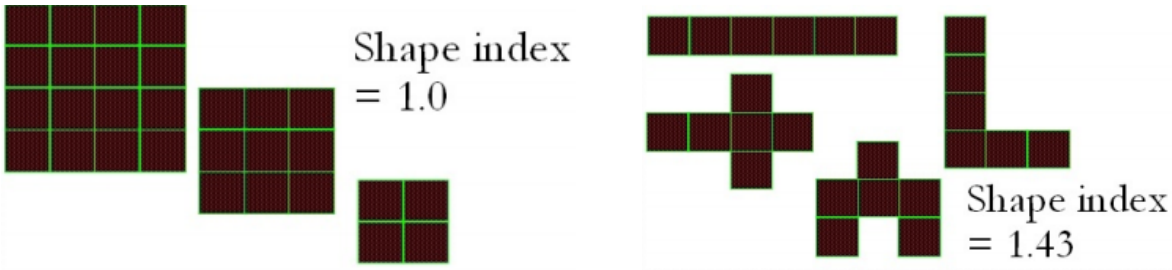


Fig. 2. 2 Example and explanation of shape index where 1 = square of any size. Adopted from McGarigal, 2017.

2.1.4 Environmental impacts on vegetation patch pattern and process

The effect of abiotic and biotic factors on vegetation composition are well documented at large (landscape) scales, but the environmental factors that affect vegetation species at intermediate- and small-scales are less understood (Bennie *et al.*, 2008). In addition to key environmental factors (e.g. soil pH, altitude, slope, soil water content and grazing) topography

also has a major role (Greig-Smith, 1979; Craine, 2005). It may affect plant patch formation whilst associated variables such as slope and aspect may also alter plant growth (Bennie *et al.*, 2008; Måren *et al.*, 2015). With the exception of grazing, such factors change relatively slowly over time, allowing plant species adapted to specific microclimates to establish (Bennie *et al.*, 2008). Soil pH and soil water content can alter plant photosynthesis or competitive ability and hence ecological dominance (Bennie *et al.*, 2008). The environment acts on individual plants at small spatial scales, altering the shapes and numbers of patches etc., and then cascade upwards to visible impacts on the emergent larger-spatial scales at field or landscape (Billings, 1952; Franklin, 1995; Bennie *et al.*, 2008).

Drainage ditches can have substantial effects on plant abundance, growth and distribution (Coulson *et al.*, 1990). The usage of drainage ditches in upland UK was introduced to blanket bogs and moorlands primarily to improve land for game birds and sheep grazing with higher productivity grassland and increased cover of *Calluna vulgaris* (heather) (Coulson *et al.*, 1990). Open, semi-natural, drainage systems may change the competitive ability of dominant and sub-dominant species, and the patches that develop. This research also tries to understand how open drainage ditches, now common in upland UK, affect dominant and subdominant patch structure. With both drainage systems and slope across a range of altitudes, water drainage can substantially vary at field-scale. Coulson *et al.* (1990) found that drainage ditches changed the vegetation nearby, typically within 10m of the ditches, but effects were lower over longer distances. This chapter addresses changes at fairly small scale, at less than 1m, and for both dominant subdominant species in terms of their small-scale patch structure (Wigmosta *et al.*, 2002).

Vegetation growth and patch patterns can also be affected by secondary influences that arise from additional physical stresses. Semi-natural upland agricultural systems form a *piosphere*, which is an area where herbivores graze and the resulting effects on soil and vegetation (Fenton, 1937; Lange, 1969). These become susceptible to herbivore trampling and intense grazing especially in regions of permanent pastures (Fenton, 1937; Lange, 1969). Trampling of vegetation by sheep and/or cattle is generally short-term but can severely reduce the number of vegetation species, abundance and vegetation height (Liddle and Greig-Smith, 1975). The trampling damage caused by herbivores results in pressure on vegetation to recuperate, grow and stabilise (Liddle and Greig-Smith, 1975; Plumptre, 1994). Plants in this region have now probably adapted to trampling effects, however, it might also mean that vegetation areas normally trampled have, over time, transformed into sheep tracks or paths.

This reduces the effect on trampling on vegetation but translates into long-term stresses on vegetation where sheep paths develop (Fenton, 1937; Plumptre, 1994).

At Ashtrees, for example, the structure and organisation of plant communities is variable and the extent of environmental pressure is not fully comprehended (Sanderson and Rushton, 1995). Whilst increased grazing pressure in the uplands can sometimes result in high vegetation diversity (Bullock and Pakeman, 1997) this does not necessarily imply greater conservation value, and indeed sometimes the reverse is true (Smith, 1996; Smith *et al.*, 2003; Smith and Wyatt, 2007). Grazing may have a disproportionate effect on dominant and subdominant species due to higher inter-specific competition and potentially a change in community classification (Adamson et al., 2001b). Herbivore grazing has also been suggested to change vegetation spatial patterns and characteristics (Watt, 1947). Sheep, for example, are generalist grazers and while they prefer plants that are more palatable (e.g. heather), they wander widely and graze within different parts of a field, even in the absence of such palatable plants.

Sheep paths develop from the navigational skills of sheep and their forage preferences (Lange, 1969). Whilst in lowland or arid areas tracks can span outwards from a watering hole, in UK uplands water is more readily available, and thus sheep paths tend to lead to palatable areas of feeding (Lange, 1969) or supplementary fodder provided by farmers (Sales *et al.*, 2016). The development of sheep paths may also cause surrounding vegetation to change (Bates, 1935) depending on distance from sheep paths, affecting vegetation patch structure and species composition (Thomas, 1959; Lange, 1969). While the density of sheep paths does not directly reflect stocking rates, their easy identification from aerial photographs can provide useful information on relative grazing pressure (Lange, 1969). Many studies have investigated the effects of herbivore grazing and trampling on vegetation but relatively few have related paths directly to vegetation patch structure at specific spatial scales (Vesk and Westoby, 2001). The study therefore asks the question of whether sheep paths affect vegetation patch dynamics at a very small scale ($1m^2$) in terms of length of sheep path around each quadrat and distance of sheep path from the set quadrat.

2.1.5 Aims and objectives

The overall aim of this chapter is to understand the biotic and abiotic factors that affect the spatial distribution of patches of both the dominant and subdominant species of vegetation at the 1m scale, and how these patterns can be scaled-up to the entire heft. The specific objectives were to:

1) compare traditional 1m quadrat-scale measures of vegetation, assessed as overall percentage cover, with those derived from 10cm-scale measures of dominant and sub-dominant species

2) relate the dominant and subdominant vegetation species to environment, particularly soil pH, hydrological influence, slope, altitude, sheep grazing and drainage

3) calculate patch metrics for vegetation at 10cm scale and determine the major environmental factors affecting these metrics

4) scale-up measures of vegetation patchiness from the 10cm and quadrat-scale to the whole Ashtrees heft.

2.2 Methodology

2.2.1 Comparison between vegetation assessed via within-quadrat and whole quadrat survey methods (Objective 1)

2.2.1.1 Collection and management of vegetation data at 10cm scale

Data for dominant and subdominant vegetation data within each $1m^2$ quadrat for 167 quadrats across the Ashtrees heft surveyed in 1991 were used as precursors for patch analysis. Each $1m^2$ quadrat was divided into 100 x 10cm grid squares with wires, and the dominant and subdominant species of vegetation estimated by eye. The original survey was undertaken by Dr. Roy A. Sanderson and Prof. Stephen P. Rushton, with records recorded on individual 167 gridded paper sheets. This data was recorded in Excel and saved as comma separated value (CSV) files such that all subsequent data management, manipulation and analysis could be undertaken in R (R Core Team, 2013). Each quadrat was initially recorded as a table of the dominant (or subdominant) species in each cell (see Table 2.1 for an example). Separate tables could then be produced for each dominant (or subdominant) species, scoring them as 0 if absent, and 1 if present in an individual 10cm cell.

Table 2. 1 Dominant species names in a $1m^2$ quadrat. This is an example of the matrix used before any manipulation and analysis was performed. Key: Nardstri = *Nardus stricta,*, Holclana = *Holcus lanatus*, Agrocapi = *Agrostis capillaris*, Anthodor = *Anthoxanthum odoratum*, Festrubr = *Festuca rubra*, Desccesp = *Deschampsia cespitosa*, Cardpalu = *Cardamine palustre*, Junceffu = *Juncus effusus*, Carenigr = *Carex nigra*, Festovin = *Festuca ovina*, Carepani = *Carex paniea*.

	10	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
10	Nardstri	Cardpalu	Festrubr	Nardstri	Festrubr	Festrubr	Festrubr	Nardstri	Nardstri	Nardstri
20	Nardstri	Nardstri	Carenigr	Holclana	Nardstri	Nardstri	Agrocapi	Nardstri	Nardstri	Nardstri
30	Nardstri	Holclana	Carenigr	Nardstri	Nardstri	Nardstri	Festovin	Nardstri	Nardstri	Nardstri
40	Holclana	Holclana	Nardstri	Festovin	Nardstri	Carenigr	Carenigr	Carenigr	Nardstri	Carenigr
50	Holclana	Holclana	Agrocapi	Festovin	Nardstri	Nardstri	Agrocapi	Carenigr	Carepani	Nardstri
60	Agrocapi	Holclana	Holclana	Nardstri	Nardstri	Nardstri	Festovin	Festrubr	Nardstri	Nardstri
70	Anthodor	Anthodor	Holclana	Nardstri	Junceffu	Nardstri	Nardstri	Nardstri	Nardstri	Nardstri
80	Festrubr	Festrubr	Festrubr	Holclana	Agrocapi	Festrubr	Nardstri	Nardstri	Nardstri	Nardstri
90	Desccesp	Nardstri	Desccesp	Anthodor	Holclana	Holclana	Holclana	Holclana	Agrocapi	Agrocapi
100	Desccesp	Desccesp	Desccesp	Festrubr	Holclana	Nardstri	Nardstri	Holclana	Desccesp	Festrubr

2.2.1.2 Multivariate comparison of sub-quadrat data with conventional survey methods

Unconstrained ordination via non-metric multidimensional scaling (NMDS) was used to summarise vegetation covers estimated for the dominant or subdominant 10cm survey. This required combining the number of records for each 10cm square for the whole 1m quadrat, producing a value of 0 to 100 for each quadrat for each dominant or subdominant species. This simply required the calculation of the sum of all 0/1 records for the incidence of each species in every 10cm cell, generating a 1m whole-quadrat scale data set. These two NMDS ordinations of the dominant and subdominant species were compared with the equivalent ordination of raw percentage cover obtained from the conventional whole-quadrat survey. 'Procrustes rotation' from the R 'vegan' package (Oksanen *et al.*, 2015) was used to compare the three ordination plots. Procrustes rotation rescales and rotates the ordinations so that they match as closely as possible (Gower, 1975; Oksanen *et al.*, 2015). 'Protest', also from the 'vegan' package (Oksanen *et al.*, 2015), was used to test the similarity m^2 (Procrustes residual derived from the sum of the squared deviation) and R^2 (correlation coefficient) of each pair of ordinations. Three comparisons were made:

i) species cover derived from percentage visual estimates versus dominant vegetation (0-100)

ii) species cover derived from percentage visual estimates versus subdominant vegetation (0-100)

iii) dominant (0-100) versus subdominant (0-100)

2.2.2 Relationship of dominant, subdominant and percent cover estimates at each quadrat with the environment (Objective 2)

2.2.2.1 Digitising sheep tracks and drainage ditches

Whilst general environmental data were available for each quadrat (soil water, pH, altitude, slope; appendix 1.1) no data were available on possible sheep grazing patterns, or locations of drainage grips, both of which might affect the vegetation. Positions of sheep paths and drainage grips were digitised in ArcGIS from a stereo-pair of aerial photographs taken in 1991 (Fig. 2.3) in which the whole farm could be clearly viewed with no obscuring by cloud cover.

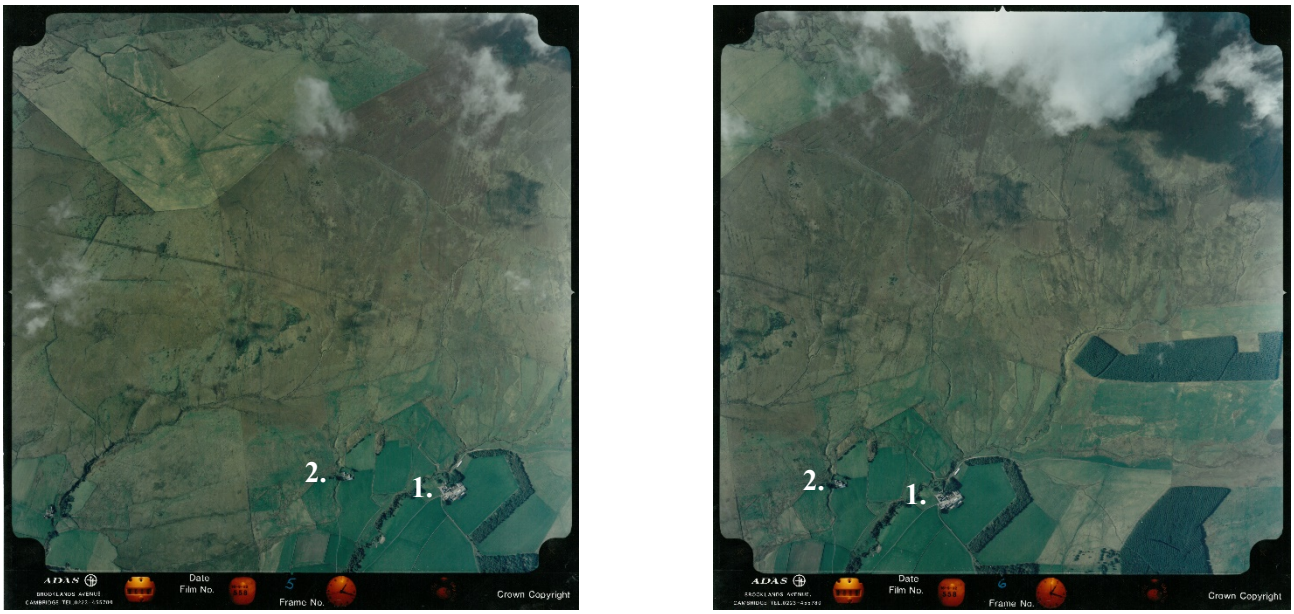


Fig. 2. 4 Aerial images used for superimposition of the Ashtrees Dipper photographed in 1991. Points 1 and 2 were main locations used for superimposition of the stereo-images.

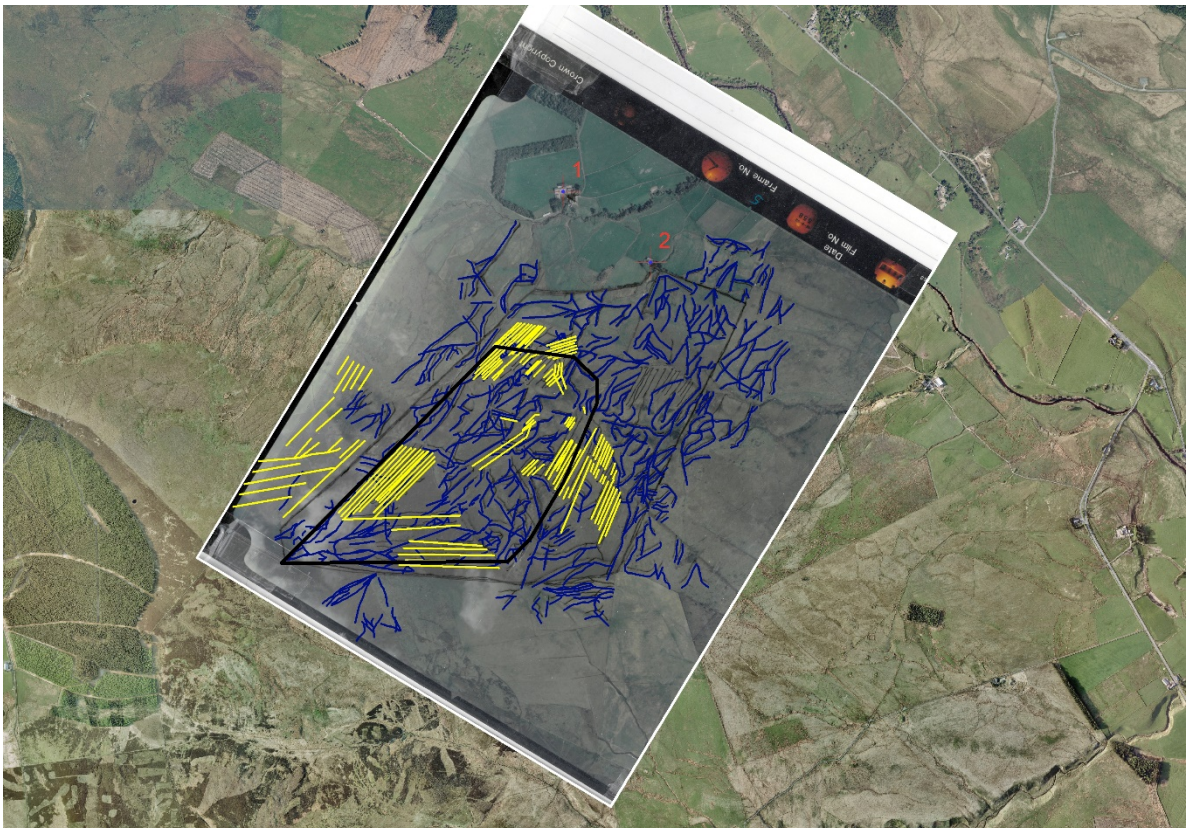


Fig. 2. 3 Hard copy aerial image digitised as a raster and sheep tracks (blue) and ditches (yellow) georectified onto OS National Grid with EDINA 2017 Digimap backdrop aerial photograph.

A stereoscope was used to view the images in 3D and high resolution. Transparency film was placed on one of the images, marking two main points of origin. The outline of the Ashtrees Dipper was marked with a 0.05 width ink pen, and both sheep tracks and drainage ditches marked onto the film. The film was scanned at high resolution, and georectified with 2017 Edina air photographic imagery to match the Ordnance Survey National Grid. Six georectification points were selected from clearly defined unchanged features (e.g. buildings, forest edges, road junctions etc.) that were unchanged between 1992 and 2017 (Fig. 2.4). Two separate maps, for sheep tracks and ditches in 1992 were the final outputs (Fig. 2.5).

Fig. 2. 5 Sheep tracks (left) and ditches (right) at Ashtrees, digitised from 1992 stereo-images using ArcMap (ESRI Inc., 2016); (single backdrop from EDINA Digimap 2018).

At each quadrat 10m, 25m and 35m buffers was generated and the total length of sheep track within each buffer calculated. The 10m buffer was selected to measure any effects of sheep tracks in close proximity to a quadrat, whilst the 35m buffer was maximum distance where no two buffers overlapped (Fig. 2.6). The total length of sheep tracks within each buffer, and distance from quadrat to nearest sheep track were calculated (Fig. 2.7). Total number of patches per quadrat, mean area of patches per quadrat and mean shape index of patches per quadrat were related to total length and distance to sheep patches via linear models. In addition, distance to nearest ditch from each quadrat was also measured.

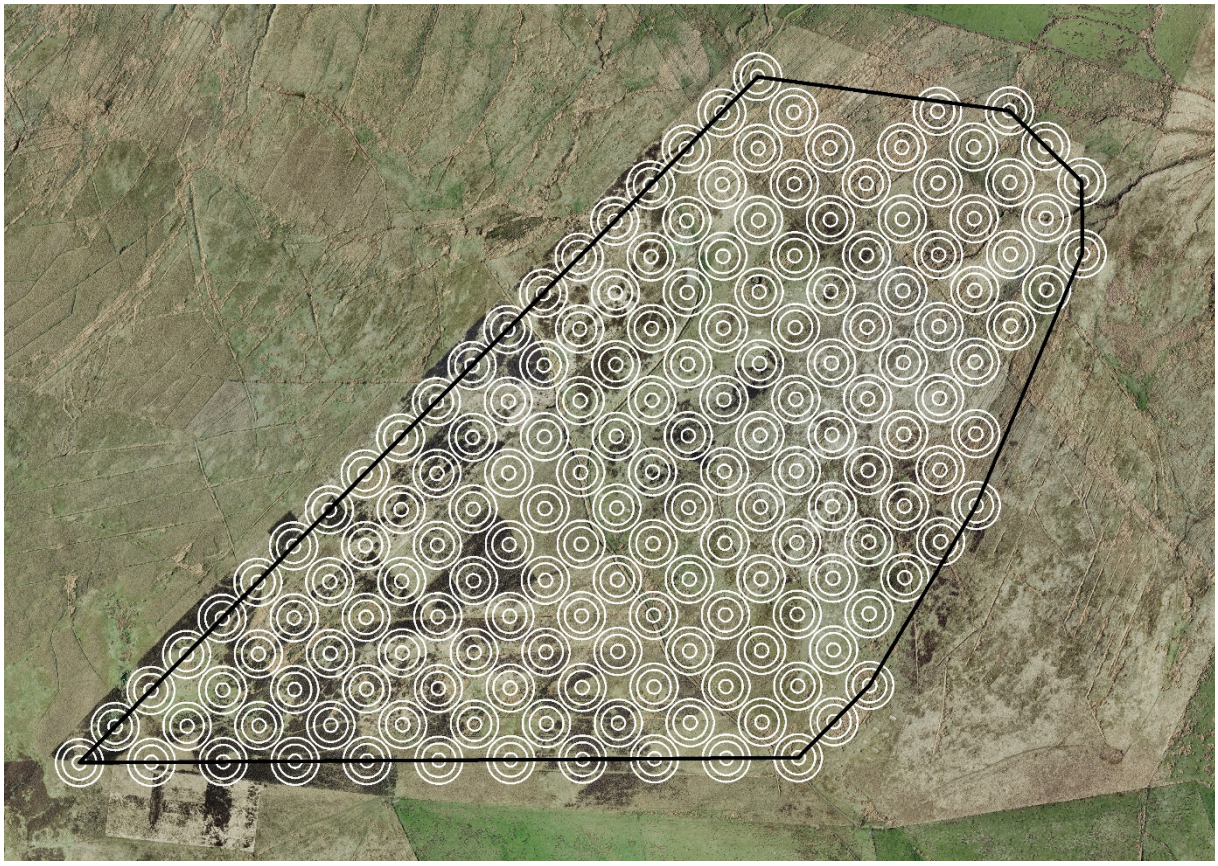


Fig. 2. 6 Buffer zones of 10m, 25m and 35m around each of 167 quadrats.

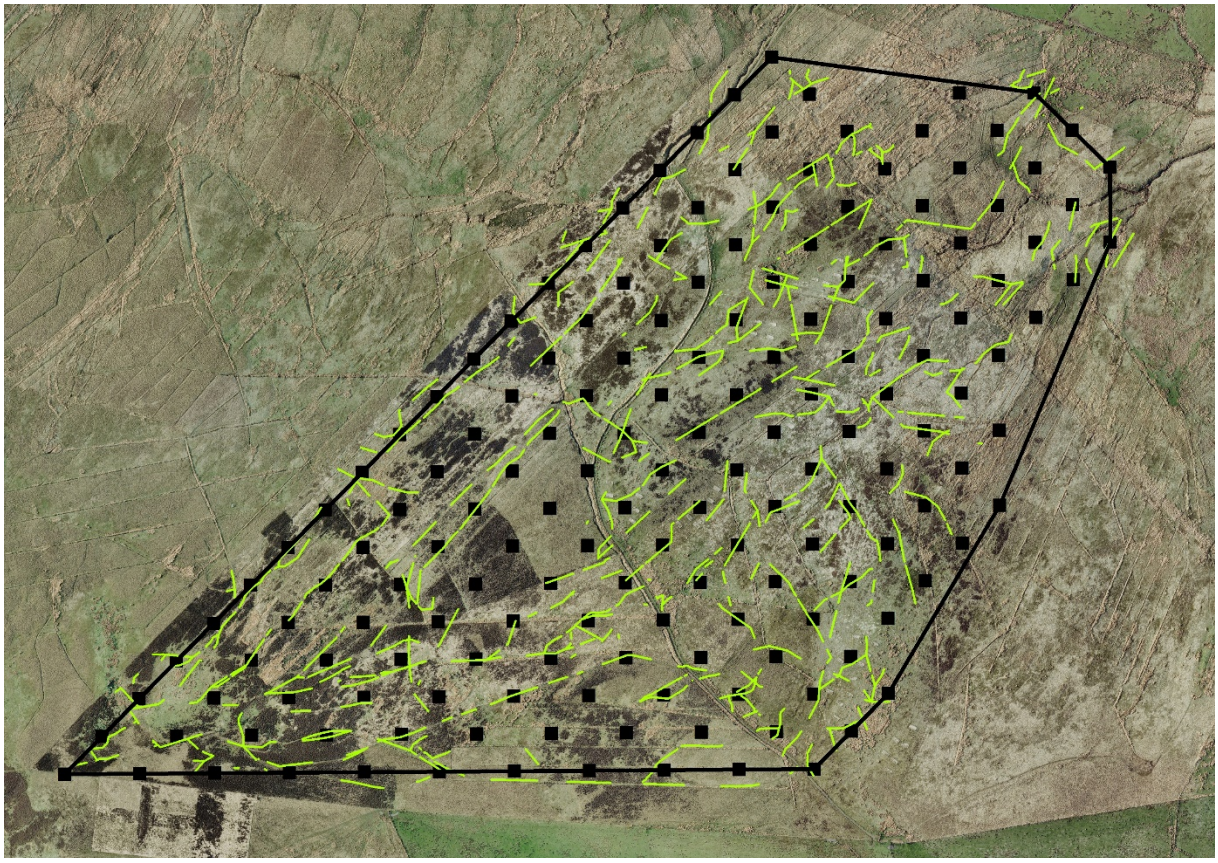


Fig. 2. 7 Intersected sheep tracks that fall within a maximum of 35m buffer zone around each quadrat.

2.2.2.2 Constrained ordination analysis of dominant, subdominant and percentage cover

Redundancy analysis (RDA), a linear constrained ordination analysis, was used to relate the three sets of vegetation community data (dominant, subdominant and percentage cover) to the environment. Canonical Correspondence Analysis (CCA), a unimodal constrained ordination method did not provide a visually representative plot of the constrained data, i.e. all the data points were clustered together in the centre of the plot (see Chapter 1 for comparison of linear and unimodal methods). Soil pH, soil water content, altitude, slope, distance to and length of sheep tracks at each of 10m, 25m and 35m buffer zones as well as distance to drainage ditches (Rodwell, 2006**;** Rushton, 1992; Sanderson *et al.*, 1995a) were used as constraining variables. Soil pH, soil water content and slope were recorded in the field whilst altitude was from EDINA Digimap (EDINA, 2018).

2.2.3 Environmental and management factors affecting the patch metrics of the quadrats (Objective 3)

2.2.3.1 Calculation of patch metrics

To extract patch metrics for a given species, a 10 x 10 matrix was created containing 1 or 0 values, where 0 represents background of other species and 1 represents presence of that particular species (example in Table 2.2). The final step was to identify different patches, encoding each patch separately (Table 2.3). Diagonally adjoining cells were assumed to belong to the same patch (Table 2.3), which can then be easily visualised by colour (Fig. 2.8).

Patch statistics were then performed on the matrices obtained for each species within each quadrat. This was done using the 'SDMTools' package available in R (VanDerWal *et al.*, 2018). The 'PatchStat' function can calculate numerous metrics, but the three of primary concern in this chapter were:

- number of patches
- area per patch
- shape index (see Introduction).

An example of the patch statistics output for a single quadrat containing 5 patches is provided in Table 2.4. A semi-automated R script was developed to input and manipulate data, extract patches and calculate patch metrics for both dominant and subdominant species at all 167 quadrats on Ashtrees Dipper.

Table 2. 2 Example of a 0-1 matrix for a given species in a given quadrat. 0 is background (i.e. species not present) and 1 represents 10 x 10 patch of the species of concern.

	10	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
10	1	1	0	0	1	0	0	0	1	1
20	1	1	1	0	0	1	1	0	1	1
30	1	1	0	0	1	1	1	0	1	1
40	0	0	0	1	0	1	0	0	0	1
50	0	1	0	0	0	1	1	0	0	0
60	0	1	0	0	1	1	1	0	0	1
70	0	1	0	0	1	0	1	1	1	1
80	0	1	0	0	0	0	0	1	1	1
90	0	0	1	0	0	0	0	0	0	0
100	0	0	0	0	0	0	1	1	0	0

Table 2. 3 Matrix of patch numbers for one species in a given quadrat. Each number (1 - 5) is a different patch formed by the species and 0 is background (i.e. species not present).

	10	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
10	1	1	0	0	2	0	0	0	3	3
20	0	0	0	0	0	0	3	3	0	0
30	4	4	4	0	0	3	3	0	3	3
40	4	4	0	0	3	3	3	0	3	3
50	0	0	0	3	0	3	0	0	0	3
60	0	5	0	0	0	3	3	0	0	0
70	0	5	0	0	3	3	3	0	0	3
8	0	5	0	0	3	0	3	3	3	3
90	0	5	0	0	0	0	0	3	3	3
100	0	0	5	0	0	0	0	0	0	0

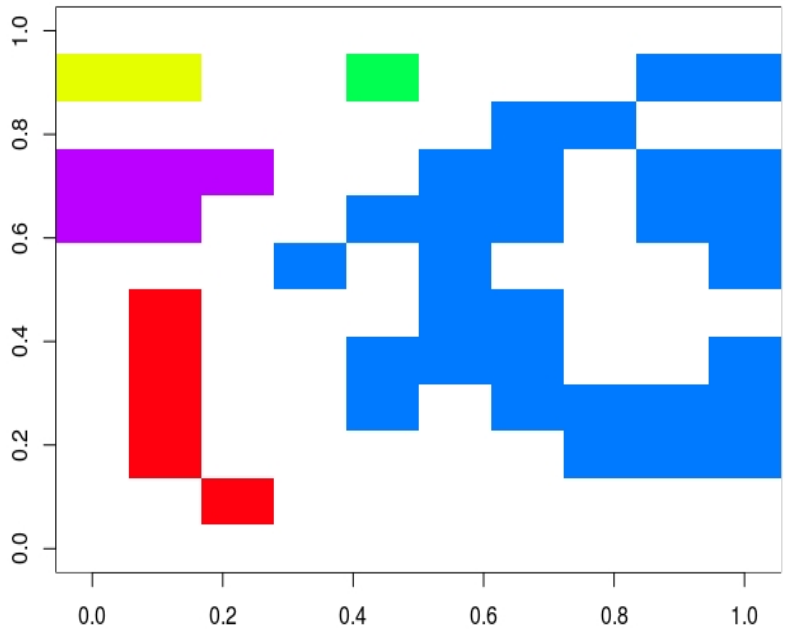


Fig. 2. 8 visual representation of the matrix in table 2.3 for a given species in a given 1m^2 quadrat. Each colour represents an independent patch of that species.

Table 2. 4 Example of patch statistics obtained for a given species within a given quadrat. 0 is background (i.e. patches where species is not present).

Patch ID	n cell	n edges perimeter	N edges internal	Area	Perim	Perim:area	Shape index	Frac dim index
0	77	116	192	77	116	1.506	3.222	1.550
1	2	6	2	2	6	3	1	1.170
2	1	4	0	1	4	4	1	NA
3	30	58	62	30	58	1.933	2.636	1.573
4	5	10	10	5	10	2	1	1.139
5	5	14	6	5	14	2.8	1.4	1.557

Shape index is an alternative to 'fractal dimension' calculated as a normalised ratio of patch perimeter to area ratio (see Introduction); it is simpler to interpret than fractal dimension which uses a log transformation (McGarigal, 2017). The patch statistics per quadrat, for 167 quadrats, were also summarised as i) total number of patches per species per quadrat, ii) mean patch area per species per quadrat and ii) mean patch shape index per species per quadrat.

2.2.3.2 Analysis of environment on overall quadrat patch metrics

Information from the scientific literature was used to categorise the most common species (in terms of incidence) into two vegetation growth strategies: 'block-former' species, which form distinct clumps, or 'opportunistic' species that occur in the background matrix but not in clumps (Robinson and Rorison, 1983; Rushton *et al.*, 1992; Smith, 1996; S. Rushton *pers comm.*). Block- formers (abbreviated names in brackets) were *Calluna vulgaris* (Callvulg), *Eriophorum vaginatum* (Eriovagi), *Juncus effusus* (Junceffu), *Juncus squarrosus* (Juncsqua), *Molinia caerulea* (Molicaer) and *Nardus stricta* (Nardstri). Opportunistic species were *Carex nigra* (Carenigr), *Deschampsia flexuosa* (Descflex), *Galium saxatile* (Galisaxa), *Potentilla erecta* (Poteerec) and *Vaccinium myrtillus* (Vaccmyrt). It should be noted that this classification is not a traditional or published classification but rather a descriptive classification that might also be site-specific since not all named block-formers or opportunistic species will display such characteristics at other locations. For example, in some situations *D. flexuosa*, *G. saxatile*, *P. erecta* and *V. myrtillus* can form extensive patches (block-formers) or grow in small tufts (opportunits) and thus might largely depend on the surrounding environment. Number of patches, patch area, and patch complexity were calculated for the two groups of species, for both dominant and subdominant records. Linear models were used to investigate the response of mean number of patches, mean patch area, and mean shape index to vegetation growth strategy and cover type (dominant or subdominant). An interaction term between vegetation growth and cover type was also included in the linear model equation. An interaction term was assessed to identify any potential dependence of vegetation growth and cover type.

2.2.3.3 Multivariate GLM analysis of patch metrics for all species in relation to environment

Three tables were created for both dominant and subdominant species (i.e. depending on their abundance on a cell-by-cell basis; note that some species can be dominant in once cell but subdominant in another cell), with each row a separate quadrat, each column a species, and the table entries being i) total number of patches ii) patch area or iii) patch shape. The 'mvabund' package in R (Wang *et al.*, 2019) was used for multivariate analysis, as this allows the appropriate error distribution to be selected for use in the 'manyglm' function, typically Poisson distribution for count response data, or Gaussian or negative binomial for continuous response data. Explanatory variables were soil pH, water content, slope, altitude, total length of sheep track, distance to sheep track and distance to ditch. These analyses were undertaken

for all three buffer distances. Note that 'manyglm' ANOVA analyses all species simultaneously, but for clarity the results only include detailed graphs for 'block-forming' and 'opportunistic' species (see below).

2.2.4 Interpolating patch metrics to the whole of Ashtrees Dipper (Objective 4)

Interpolation was used to create continuous predicted maps across the whole of Ashtrees Dipper of the patch metrics, rather than just at the individual observation quadrat points. Inverse distance weighted (IDW) interpolation was used as it avoided interpolation to negative values across the heft, producing a minimum value of zero for number of patches and area per patch for each species. Other interpolation methods such as kriging and kernel smoothing gave errors such as negative numbers of patches on some parts of the interpolated map. Similar interpolation methods were used for selected species to produce maps of i) number of patches, ii) mean area of patches and iii) mean shape index across the heft. All these maps were used to visually aid interpretation of the quantitative analyses already described above.

2.3 Results

2.3.1 Comparison between vegetation assessed via within-quadrat and whole quadrat survey methods (Objective 1)

The species NMDS plots for conventional percentage cover estimates, dominant and subdominant are shown in Figs. 2.9 to Fig. 2.11 respectively, and there were similarities between the methods, with species *Sphagnum* species (e.g. *S. capillifolium* in Figs. 2.9 and 2.11) present in being at the extremes of NMDS Axis 1, *Rumex* species (mostly *R. acetosa* in Figs. 2.10 and 2.11), and *Holcus mollis* (Figs. 2.10 and 2.11) at the lower extremes of Axis 2. Formal comparison between the methods was via Procrustes rotation of the samples (167 quadrats), and summarised in Table 2.5.

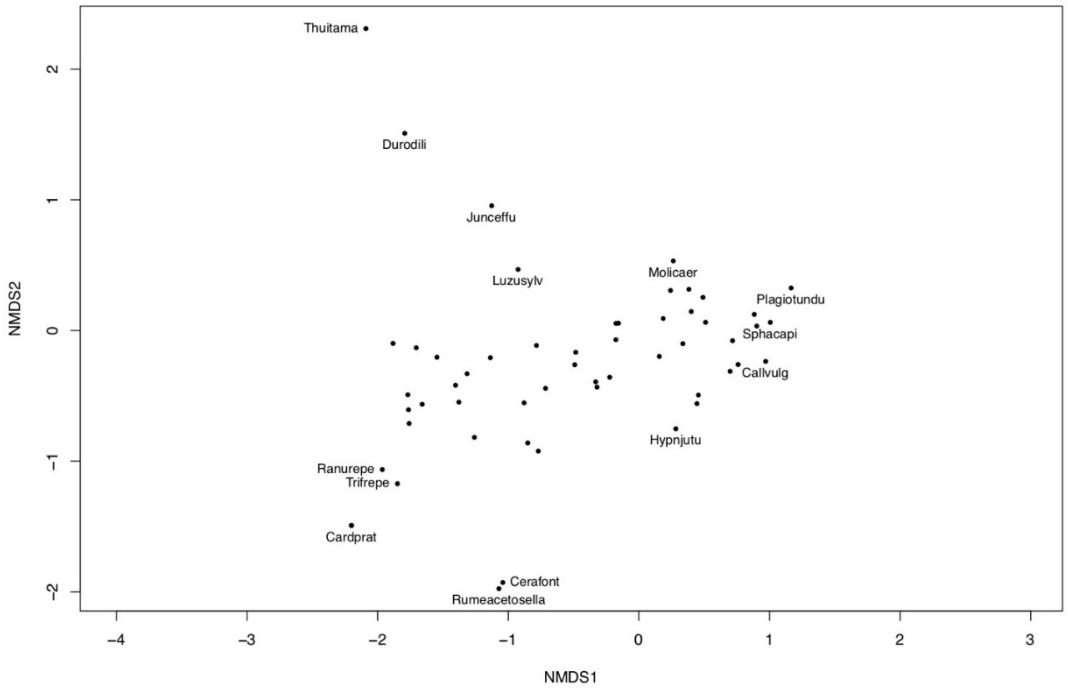


Fig. 2. 9 NMDS ordination species plot from percentage cover abundance data at Ashtrees. Thuitama = *Thuidium tamariscinum*, Durodili = *Dryopteris dilitata*, Juneffu = *Juncus effusus*, Luzusylv = *Luzula sylvatica,* Molicaer = *Molinia caerulea*, Plagiotundu = *Plagiothecuim undulatum*, Sphacapi = *Sphagnum capillofolium,* Callvulg = *Calluna vulgaris*, Hypnjutu = *Hypnum jutlandicum*, Cetafont = *Cerastium fontanum*, Rumeacetosella = *Rumexacetosella*, Cardprat = *Cardemine pratensis*, Trifrepe = *Trifolium repens*, Ranurepe = *Ranunculus repens*

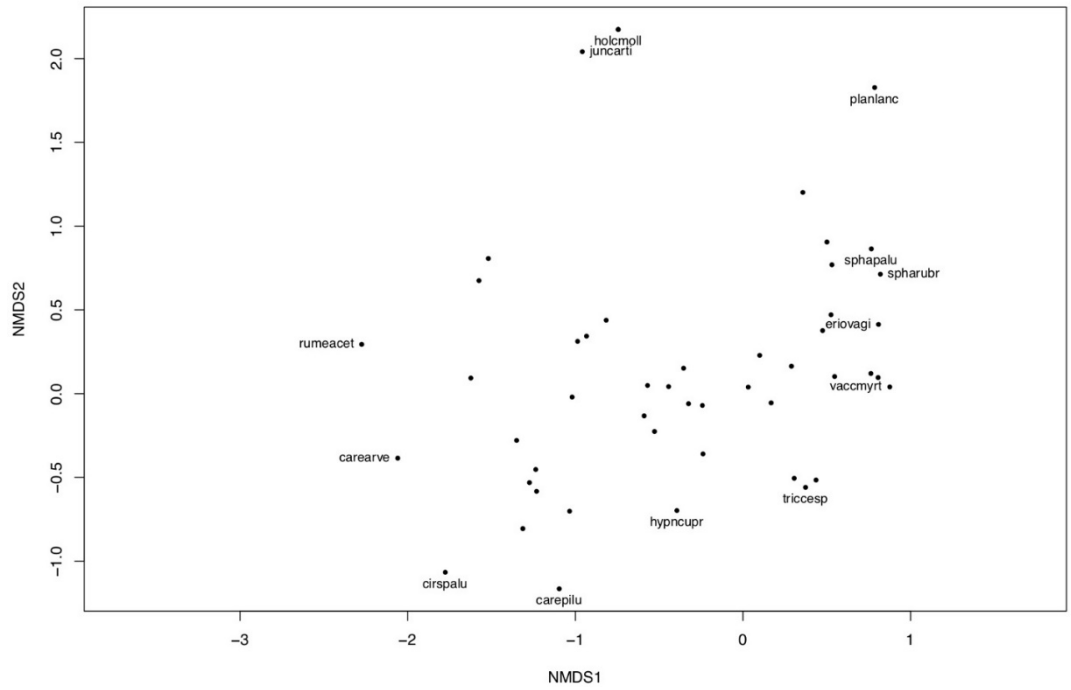


Fig. 2. 11 NMDS ordination species plot for dominant vegetation species found at Ashtrees. Holcmolli = *Holcus lanatus*, planlanc = *Plantego lanceolata*, sphapalu = *Sphagnum palustre*, spharubr = *Sphagnum rubra*, eriovagi = *Eriophorum vaginatum*, vaccmyrt = *Vaccinium myrtillus*, triccesp = *trichophorum cespitosum*, hypncupr = *Hypnum cupressiforme*, carepilu = *Carex pilulifera,* cirspalu = *Cirsium palustre*, Carearve = *Cirsium arvense*, rumeacet = *Rumex acetosa*

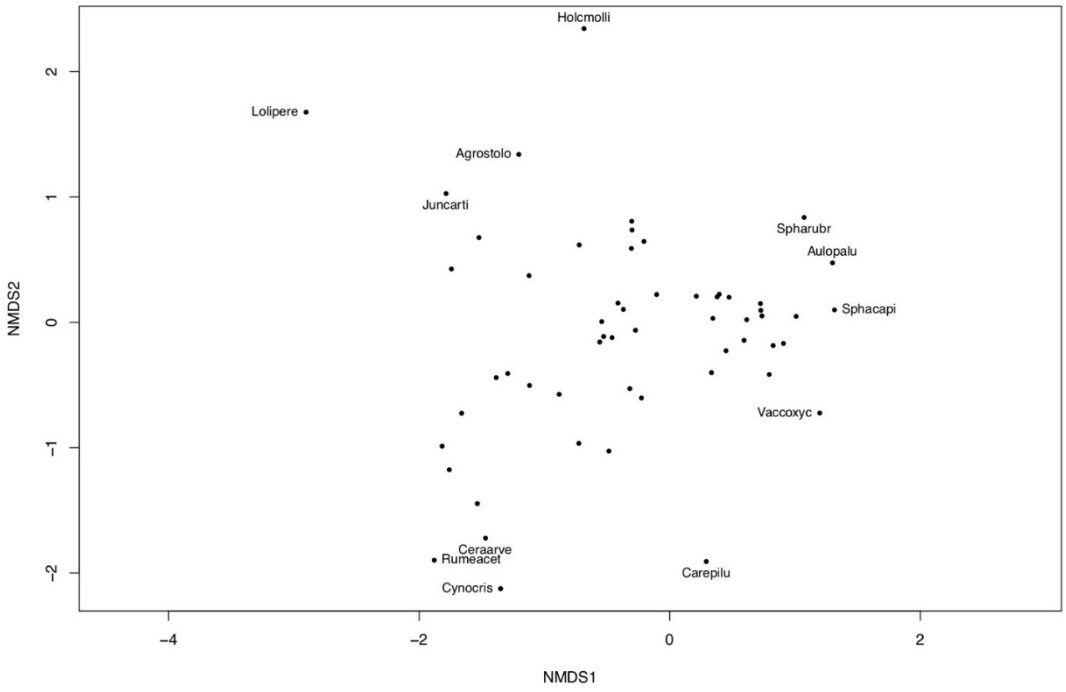


Fig. 2. 10 NMDS ordination species plot for subdominant vegetation found at Ashtrees. Holcmolli = *Holcus lanatus*, Spharubr = *Sphagnum rubra*, Aulopalu = *Aulocomnium palustre*, Sphacapi = *Sphagnum capillofolium*, Vaccoxyc = *Vaccinium oxycoccos*, Carepilu = *Carex pilulifera*, Cynocris = *Cynosurus cristatus*, Rumeacet = *Rumex acetosa*, Ceraarve = *Cirsium arvense*, Juncarti = *Juncus articulatus*, Agrostolo = *Agrostis stolonifera*, Lolipere = *Lolium perenne*

Table 2. 5 Summary of Procrustes rotation comparison between i) percentage cover data, denoted as 'survey', and dominant patch area data, ii) survey versus subdominant vegetation patch area data and iii) dominant vs subdominant vegetation patch area data. 'm^2' is Procrustes residual, R = correlation coefficient

Procrustes rotation comparison	m^2	R	*p*-value
Survey vs Dominant	0.5565	0.6659	0.001
Survey vs Subdominant	0.5111	0.6992	0.001
Dominant vs Subdominant	0.5255	0.6888	0.001

The Procrustes residual (m^2) ranges between 0.51 to 0.56 and R at about 67%-70% for all three ordination comparisons (table 2.5). The highest similarity is observed between percentage abundance and subdominant vegetation ordinations and an R of 0.6992. Lowest correlation was between percentage abundance and dominant vegetation patch area. Most of the similarity appeared to be on NMDS 1, which might be expected as this will capture a greater amount of variability than NMDS 2. The NMDS scores on the first axis were compared via linear models and were also fit to compare the extent of correlation between the three ordination sets (Table 2.6 and Fig. 2.12). Overall there was a high correlation between the axes for all three methods, particularly the dominant versus subdominant.

Table 2. 6 R^2 coefficient of determination and p-values for results from linear correlation between non-metric multidimensional scaling axes 1 (NMDS1) between, i) dominant and surveyed data, ii) subdominant and surveyed data and iii) dominant and subdominant data

NMDS axes comparison type	R^2	p-value
NMDS 1 surveyed vs NMDS 1 dominant	0.6542	<0.0001
NMDS 1 surveyed vs NMDS 1 subdominant	0.6147	<0.0001
NMDS 1 dominant vs NMDS 1 subdominant	0.7351	<0.0001

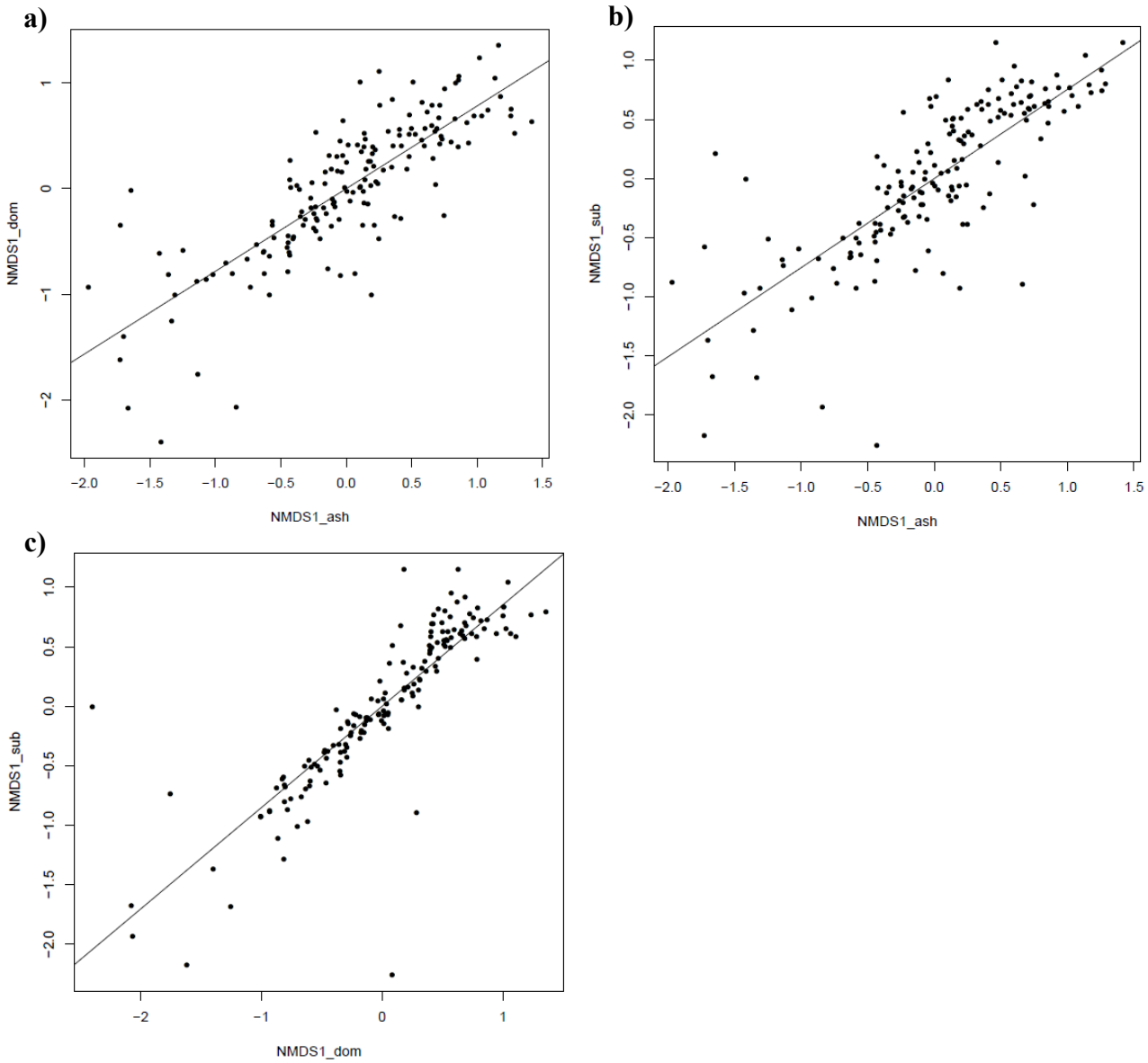


Fig. 2. 12 Comparison of non-metric multidimensional scaling results for axis 1 (NMDS1) between, a) dominant and surveyed data, b) subdominant and surveyed data and c) dominant and subdominant data.

2.3.2 Relationship of dominant, subdominant and percent cover estimates at each quadrat with the environment (Objective 2)

RDA results for percentage cover, dominant and subdominant data is presented in table 2.7 with soil pH and soil % water showing significance across the three vegetation survey types; slope shows significance for dominant and subdominant while altitude shows significance for traditional % abundance survey. Sheep tracks within a 10m buffer only shows significance with subdominant vegetation data survey. Percentage cover data shows some relationships with altitude and soil water content (Fig. 2.13). Soil pH, slope, proximity of sheep tracks and ditches result in no significance. Dominant and subdominant patch area data was used in RDA

because it is the closest numerical comparative to percentage cover data. Fig. 2.14 shows that altitude has the largest effect on dominant vegetation patch area followed by soil water content and pH. The length and direction of the arrows in the RDA 'biplots' indicate that soil pH appeared to be a major factor affecting subdominant vegetation patch structure, with altitude potentially having a less of an effect (Fig. 2.15). It is important to note that this is a visual estimation of the results of the RDA and not estimating the size of the effect of the variables based on the resulting *p*-values. ANOVA of RDA outputs was performed resulting in an overall *p*-value of 0.001. Soil pH, altitude and soil water content also result in $p = 0.001$ and slope had a significance of $p = 0.028$. Proximity (length and distance) to sheep track within 10m, 25m and 35m buffer zones show no significance. Overall ANOVA for subdominant vegetation patch area shows significance for soil pH, slope and water content. Altitude, proximity to sheep tracks and proximity to ditch show no significance with subdominant vegetation patch area. Relationships between some environmental variables were as might have been expected, for example slope and percentage water tended to be negatively correlated (arrows on RDA plots point in opposite directions).

Table 2. 7 RDA results (p- and F-) for % abundance, dominant and subdominant vegetation data and the constraining environmental variables

	% Abundance		**Dominant**		**Subdominant**	
Environmental driver	**F**	***p***	**F**	***p***	**F**	***p***
Soil pH	**6.111**	**0.001**	**3.952**	**0.004**	**5.282**	**0.001**
Slope	0.860	0.473	**2.446**	**0.015**	**2.769**	**0.007**
% water	**5.643**	**0.001**	**3.478**	**0.004**	**1.176**	**0.001**
Altitude	**5.736**	**0.002**	1.521	0.151	3.334	0.294
ST length (10m buffer)	2.312	0.062	0.737	0.605	**1.977**	**0.038**
ST distance (10m buffer)	1.651	0.142	0.687	0.686	0.673	0.760
ST length (25m buffer)	0.378	0.862	0.473	0.892	0.770	0.641
ST distance (25m buffer)	0.230	0.058	0.584	0.784	0.923	0.482
ST length (35m buffer)	0.931	0.453	0.822	0.550	0.354	0.979
ST distance (35m buffer)	1.694	0.161	1.129	0.334	0.612	0.786
Ditch distance	1.694	0.120	1.577	0.130	0.647	0.772

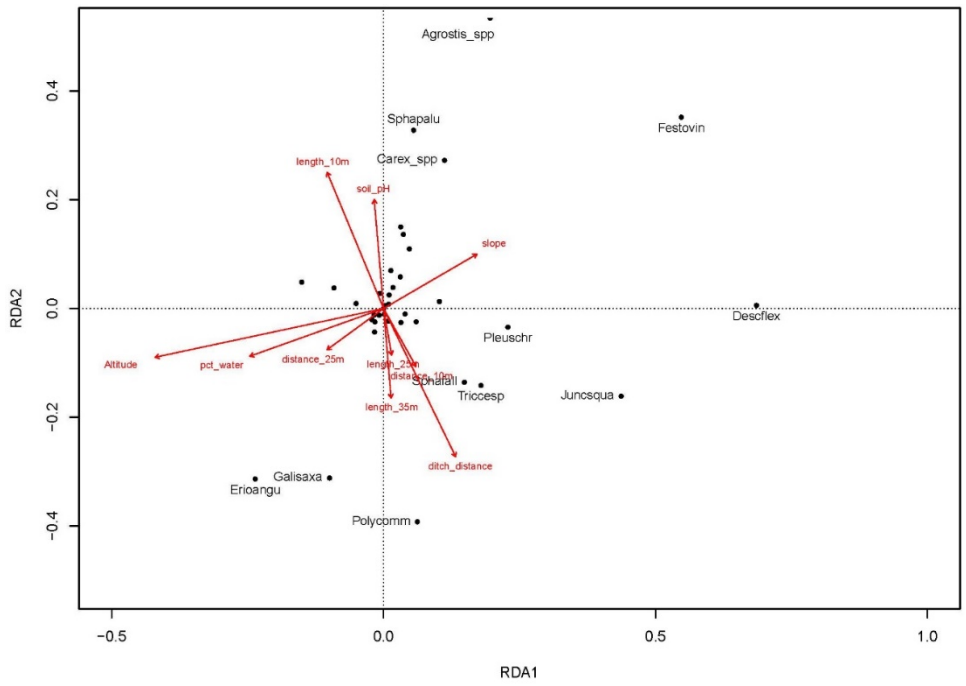


Fig. 2. 14 RDA of vegetation % abundance, with soil pH, altitude, slope, soil water content, sheep track proximity within 3 different buffer zones and distance from drainage ditch as the main constraints. Festovin = *Festuca ovina*, Descflex = *Deschampsia flexuosa*, Pleuschr = *Pleurozium schreberi*, Juncsqua = *Juncus squarrosus*, Triccesp = *Trichophorum cespitosum*, Sphafall = *Sphagnum fallax*, Polycomm = *Polytrichum commune*, Galisaxa = *Galium saxatile*, Erioangu = *Eriophorum angustifolium, Sphapalu = Spahgnum palustre*

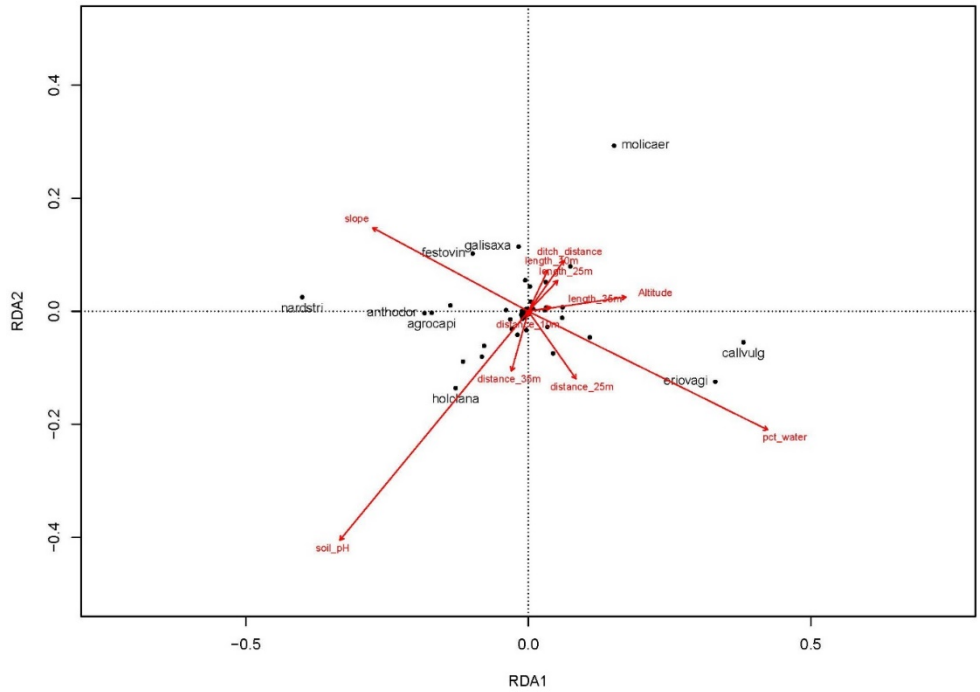


Fig. 2. 13 RDA of dominant vegetation, with soil pH, altitude, slope, soil water content, sheep track proximity within 3 different buffer zones and distance from drainage ditch as the main constraints. Molicaer = *Molinia caerulea*, Callvulg = *Calluna vulgaris*, Eriovagi = *Eriophorum vaginatum*, Holclana = *Holcus lanatus*, Agrocapi = *Agrostis capillaris*, anthodor = *Anthoxanthum odoratum*, Nardstri = *Nardus stricta*, Festovin = *Festuca ovina*, Galisaxa = *Galium saxatile*

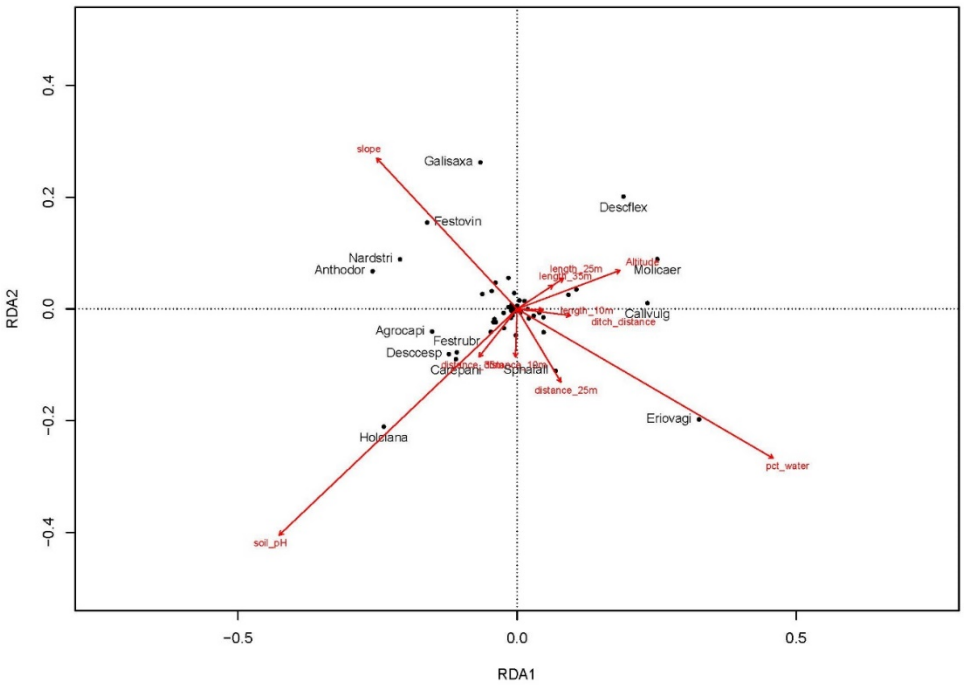


Fig. 2. 15 RDA of subdominant vegetation, with soil pH, altitude, slope, soil water content, sheep track proximity within 3 different buffer zones and distance from drainage ditch as the main constraints. Galisaxa = *Galium saxatile*, Descflex = *Deschampsia flexuosa*, Molicaer = *Molinia caerulea*, Callvulg = *Calluna vulgaris*, Eriovagi = *Eriophorum vaginatum*, Sphafall = *Sphagnum fallax*, Carepani = *Carex panacea*, Holclana = *Holcus lanatus*, Desccesp = *Deschampsia cespitosa*, Festrubr = *Festuca rubra*, Agrocapi = *Agrostis capillaris*, Anthodor = *Anthoxanthum odoratum*, Nardstri = *Nardus stricta*, Festovin = *Festuca ovina*.

2.3.3 Environmental and management factors affecting the patch metrics of the quadrats (Objective 3)

2.3.3.1 Overall patch metrics

Overall number of patches, patch area and shape complexity in block-forming and opportunistic vegetation types for the dominant and subdominant covers are shown in Figs 2.16 to Fig. 2.18. Number of patches was similar between block-forming and opportunistic vegetation types (Fig. 2.16: $F_{1,18} = 0.805$, $p = 0.987$). The mean number of patches was significantly higher for subdominant than dominant vegetation ($F_{1,18} = 13.108$, $p = 0.002$), but there was no significant interaction between vegetation growth strategy and cover ($F_{1,18} = 0.084$, $p = 0.776$). There were major differences in the mean area of patches for both the main effects and interaction terms (Fig 2.17). The area of patches was greater for the block-formers than opportunists ($F_{1,18} = 9.737$, $p < 0.001$), and larger for the dominants than the subdominants ($F_{1,18} = 7.858$, $p = 0.006$). The interaction term indicated that mean patch area

was greater in dominant block-former species, and in sub-dominant opportunistic species ($F_{1,18} = 10.983$, p=0.004). All three predictors also significantly affected shape complexity. Block formers had more complex shapes than opportunists ($F_{1,18} = 13.654$, $p < 0.001$) and there was a significant interaction term. In block-formers, the dominant species had the most complex patch shapes, whereas amongst opportunist species the subdominants had the most complex shape (Fig. 2.18, $F_{1,18} = 12.386$, $p = 0.002$).

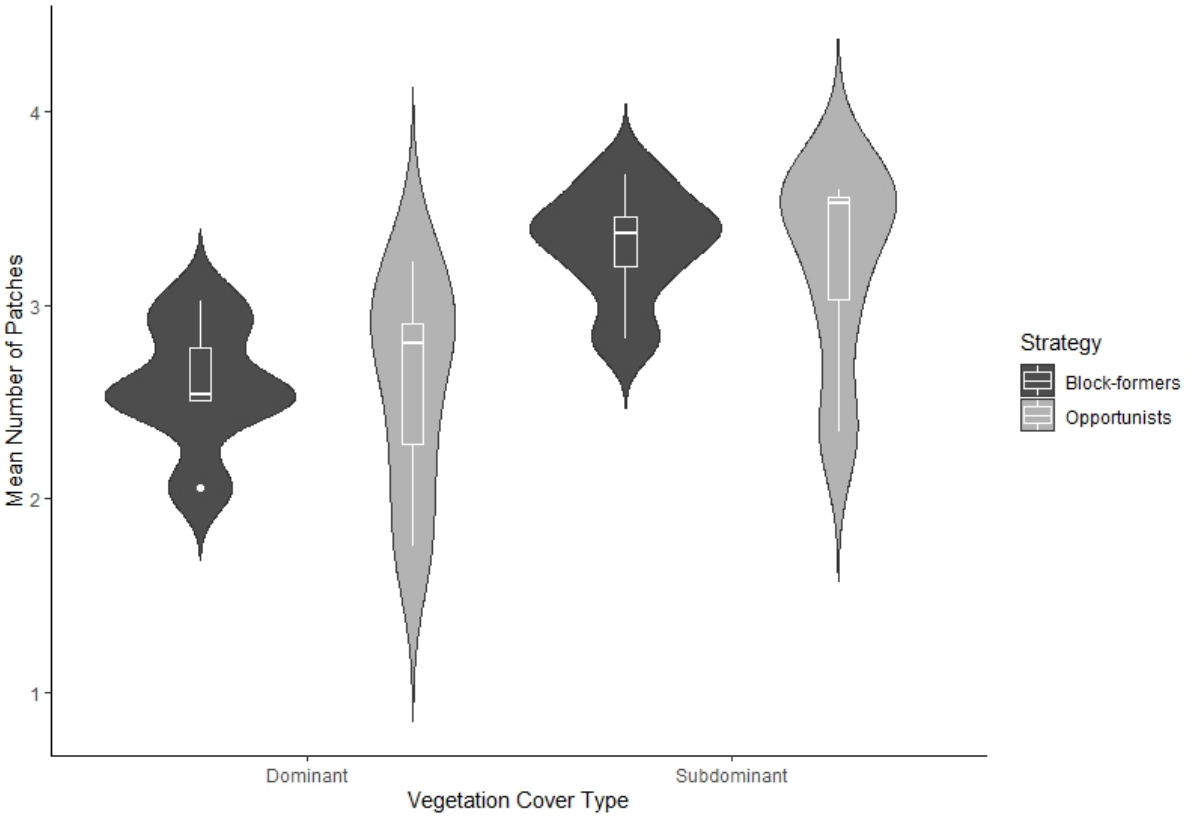


Fig. 2. 16 Violin and box-whiskers plot comparing number of patches between block-forming and opportunistic dominant and subdominant vegetation. Vegetation Cover type refers to 'Dominant' and 'Subdominant' and 'Strategy refers to 'block-formers' and 'opportunists'.

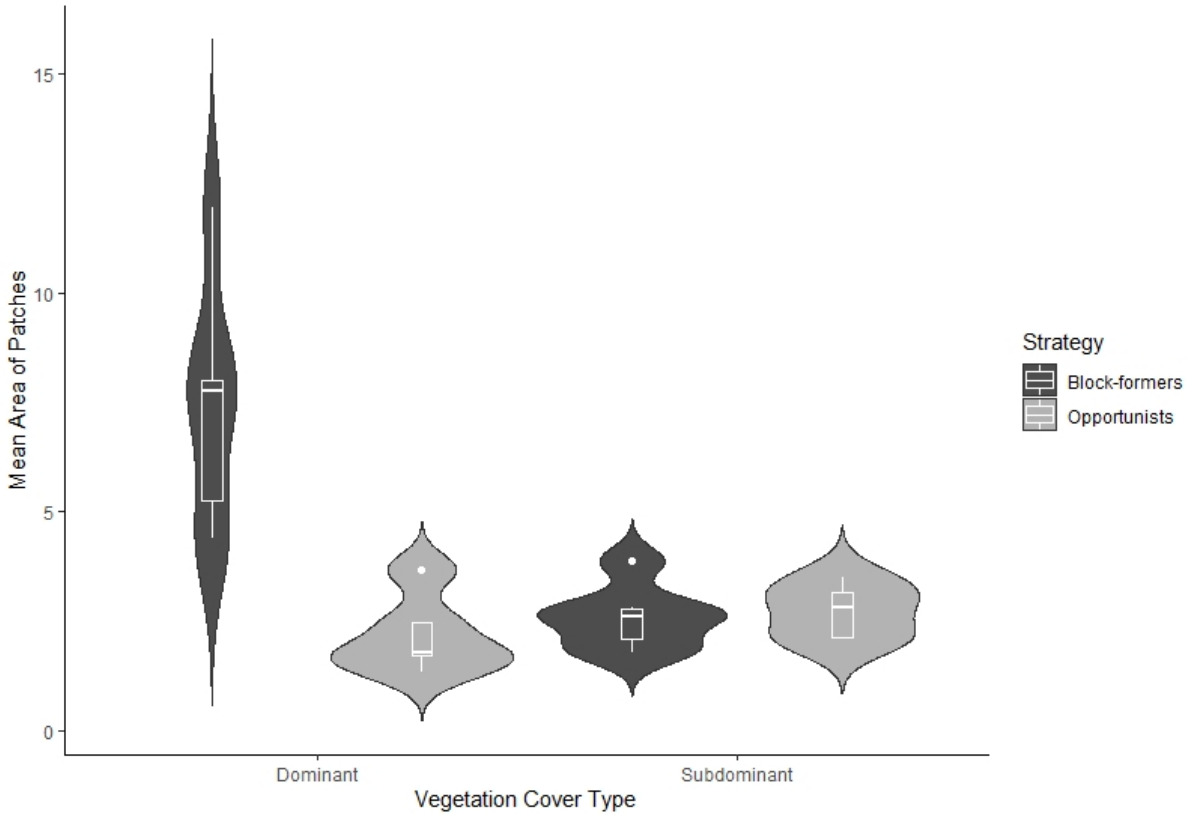


Fig. 2. 17 Violin and box-whiskers plot comparing area of patches between block-forming and opportunistic dominant and subdominant vegetation. Vegetation Cover type refers to 'Dominant' and 'Subdominant' and 'Strategy refers to 'block-formers' and 'opportunists'.

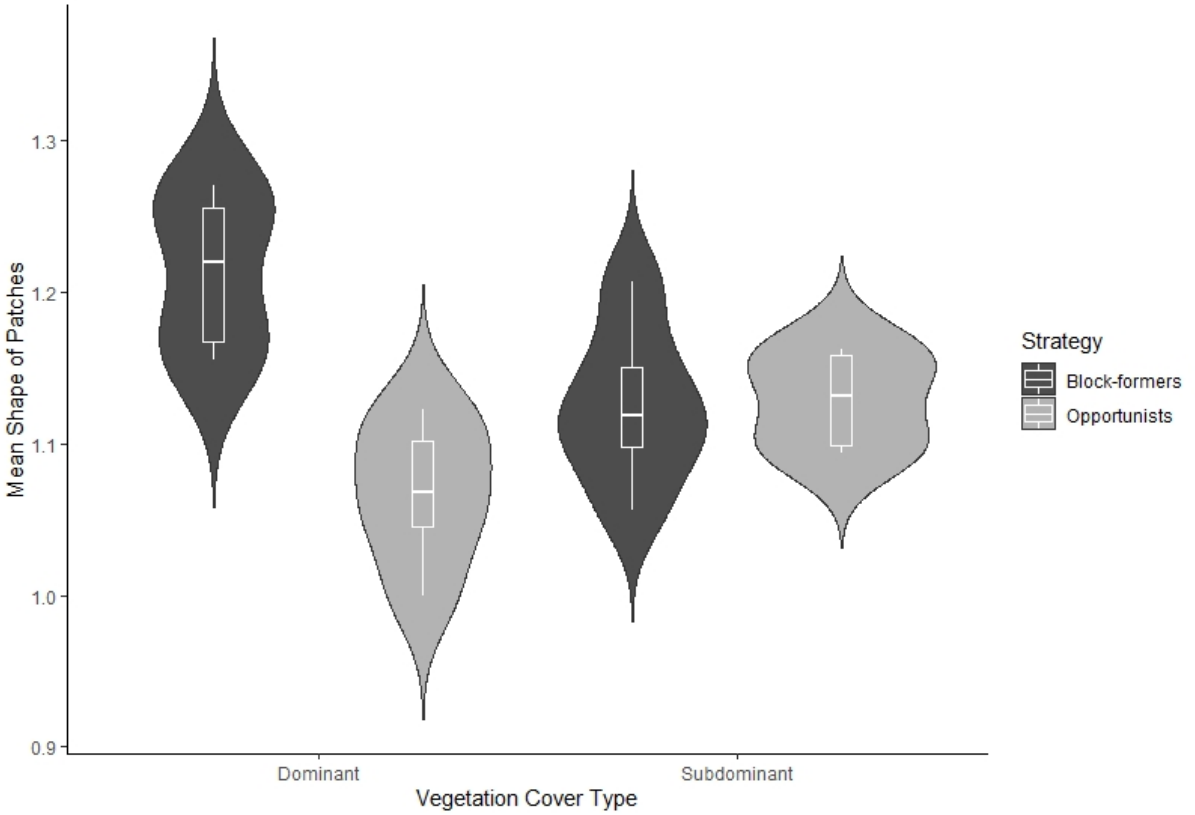


Fig. 2. 18 Violin and box-whiskers plot comparing shape of patches between block-forming and opportunistic dominant and subdominant vegetation. Vegetation type refers to 'Dominant' and 'Subdominant' and 'Condition' refers to 'block-formers' and 'opportunists'

2.3.3.2 Multivariate GLM analysis of patch metrics for all species in relation to the environment

Results of the multivariate GLM analyses with the environmental variables are summarised in Tables 2.8-2.10 (dominant) and 2.11-2.13 (subdominant). Note that a large number of *p*-statistics are presented in these tables, but each table presents the results of a single multivariate GLM analysis (via the manyglm function), thus reducing multiple testing errors, most notable Type I errors (Wang et al., 2019). The 'manyglm' approach is reported to improve the power across the range of species with different variances (Wang et al., 2019). This method also includes an assumption that there is a mean-variance relationship, fitting a generalised linear model (GLM) to each response variable and resamples the data to test for significance in the species response to the environmental data (Blakey and Letten, 2016).

Overall analysis of dominant vegetation, including less frequent species than the top eleven block-formers and opportunists, indicated strong associations with the four main environmental variables of soil pH, slope, percentage water and altitude (Tables 2.8 to 2.10). However, relatively few of the selected block formers and opportunists showed strong associations with these predictors. Amongst the block formers *Eriophorum vaginatum* was affected most, responding (both positively and negatively) to slope, and soil water content for number of patches, patch area and shape index. *Juncus effusus*, also a block-former, showed reduced shape index complexity at higher altitude. Amongst dominant opportunist species, patch area of *Deschampsia flexuosa* was positively associated with soil water content. Again, for the subdominant vegetation, overall analysis indicated that all three patch metrics were affected (positively and negatively) by the four environmental variables (Tables 2.11 to 2.13). However, no consistent patterns were apparent amongst the most common eleven species, although relationships were detected for *Nardus stricta, Molinia caerulea* (block formers) and *Deschampsia flexuosa* (opportunist).

The equivalent responses of the vegetation in relation to sheep tracks and drainage ditches are summarised in Tables 2.14-2.16 (dominant) and Tables 2.17-2.19 (subdominant). Here, even for the overall analyses that included rarer species, there were no obvious patterns, although there was evidence that proximity of sheep tracks, especially within 10m of a quadrat, affected the patch structure, and (weaker) evidence that distance to the nearest drainage ditch also had an effect. None of the most common block-forming or opportunist species showed any significant response individually to these predictors.

Table 2. 8 ManyGLM analysis of number of patches created by dominant vegetation and significance of interaction with primary environmental factors. Green = positive correlation, red = negative correlation. (p-values)

		Block Formers						Opportunists				
Env. variable	**Overall**	**Callvulg**	**Nardstri**	**Molicaer**	**Eriovagi**	**Juncsqua**	**Junceffu**	**Carenigr**	**Galisaxa**	**Poteerec**	**Descflex**	**Vaccmyrt**
Soil pH	0.001	0.424	0.999	0.421	0.591	0.941	0.999	0.993	0.967	0.999	0.135	0.889
Slope	0.001	1.000	0.044	1.000	0.003	0.991	1.000	1.000	0.112	0.999	1.000	1.000
% water	0.001	0.999	0.998	1.000	0.001	0.999	0.998	1.000	1.000	0.999	1.000	0.999
Altitude	0.001	0.966	1.000	1.000	1.000	1.000	0.075	1.000	0.985	0.899	1.000	1.000

Table 2. 9 ManyGLM analysis of area occupied by dominant vegetation and significance of interaction with primary environmental factors. Green = positive correlation, red = negative correlation. (p-values)

		Block Formers						Opportunists				
Env. variable	**Overall**	**Callvulg**	**Nardstri**	**Molicaer**	**Eriovagi**	**Juncsqua**	**Junceffu**	**Carenigr**	**Galisaxa**	**Poteerec**	**Descflex**	**Vaccmyrt**
Soil pH	0.001	0.880	0.969	0.281	0.880	0.969	0.997	0.997	0.988	1.000	0.443	0.816
Slope	0.069	0.987	0.981	1.000	0.998	0.997	1.00	1.000	0.999	0.998	0.998	1.000
% water	0.029	0.934	0.960	1.000	0.039	0.932	1.00	1.000	1.000	0.977	0.039	1.000
Altitude	0.053	0.703	0.998	1.000	0.998	1.000	0.122	1.000	1.000	0.998	0.988	1.000

Table 2. 10 ManyGLM analysis of shape index for dominant vegetation and significance of interaction with primary environmental factors. Green = positive correlation, red = negative correlation. (p-values)

		Block Formers						Opportunists				
Env. variable	**Overall**	**Callvulg**	**Nardstri**	**Molicaer**	**Eriovagi**	**Juncsqua**	**Junceffu**	**Carenigr**	**Galisaxa**	**Poteerec**	**Descflex**	**Vaccmyrt**
Soil pH	0.001	1.000	1.000	1.000	0.842	1.000	0.974	0.992	0.720	1.000	0.072	0.914
Slope	0.001	1.000	0.417	1.000	0.022	0.995	1.000	1.000	0.736	1.000	1.000	1.000
% water	0.001	1.000	1.000	1.000	0.001	0.933	1.000	0.563	0.939	1.000	0.997	0.935
Altitude	0.048	1.000	1.000	1.000	1.000	1.000	0.016	1.000	1.000	0.929	1.000	1.000

Table 2. 11 ManyGLM analysis of number of patches created by subdominant vegetation and significance of interaction with primary environmental factors. Green = positive correlation, red = negative correlation. (p-values)

		Block Formers						Opportunists				
Env. variable	**Overall**	**Callvulg**	**Nardstri**	**Molicaer**	**Eriovagi**	**Juncsqua**	**Junceffu**	**Carenigr**	**Galisaxa**	**Poteerec**	**Descflex**	**Vaccmyrt**
Soil pH	0.001	0.258	1.000	0.043	0.854	1.000	1.000	1.000	0.996	1.000	0.003	0.720
Slope	0.019	1.000	0.842	1.000	0.891	0.976	1.000	1.000	0.805	1.000	0.998	1.000
% water	0.002	0.526	0.672	1.000	0.005	0.999	0.969	1.000	0.808	1.000	1.000	0.977
Altitude	0.011	0.423	1.000	1.000	1.000	1.000	0.116	1.000	1.000	0.997	0.999	1.000

Table 2. 12 ManyGLM analysis of area occupied by subdominant vegetation and significance of interaction with primary environmental factors. Green = positive correlation, red = negative correlation. (p-values)

		Block Formers						Opportunists				
Env. variable	**Overall**	**Callvulg**	**Nardstri**	**Molicaer**	**Eriovagi**	**Juncsqua**	**Junceffu**	**Carenigr**	**Galisaxa**	**Poteerec**	**Descflex**	**Vaccmyrt**
Soil pH	0.001	0.178	0.999	0.574	0.919	0.984	1.000	0.993	0.991	1.000	0.014	0.990
Slope	0.036	1.000	0.957	1.000	0.975	0.996	1.000	0.858	1.000	1.000	1.000	1.000
% water	0.033	0.996	0.986	0.972	0.064	0.997	1.000	1.000	1.000	1.000	1.000	0.993
Altitude	0.066	0.361	0.996	0.999	1.000	0.999	1.000	0.668	1.000	0.955	1.000	1.000

Table 2. 13 ManyGLM analysis of shape index for subdominant vegetation and significance of interaction with primary environmental factors. Green = positive correlation, red = negative correlation. (p-values)

		Block Formers						Opportunists				
Env. variable	**Overall**	**Callvulg**	**Nardstri**	**Molicaer**	**Eriovagi**	**Juncsqua**	**Junceffu**	**Carenigr**	**Galisaxa**	**Poteerec**	**Descflex**	**Vaccmyrt**
Soil pH	0.001	0.414	0.991	1.000	0.878	1.000	1.000	1.000	0.873	1.000	1.000	1.000
Slope	0.001	1.000	0.102	1.000	0.102	0.950	1.000	1.000	0.164	1.000	1.000	0.983
% water	0.001	0.256	0.033	1.000	0.001	0.978	0.506	1.000	0.881	1.000	1.000	0.986
Altitude	0.001	0.973	0.983	1.000	1.000	1.000	0.256	1.000	1.000	0.997	1.000	1.000

Table 2. 14 ManyGLM analysis of number of patches for dominant vegetation and correlation with length of sheep track and distance to sheep track within a 10m, 25m and 35m buffer and distance to nearest ditch. ST = Sheep Track. Green = positive correlation, red = negative correlation. (p-values)

		Block-formers						Opportunists				
No. of Patches	**overall**	**Callvulg**	**Eriovagi**	**Junceffu**	**Juncsqua**	**Molicaer**	**Nardstri**	**Carenigr**	**Descflex**	**Galisaxa**	**Poteerec**	**Vaccmyrt**
ST length (10m buffer)	0.720	0.998	1.000	0.998	1.000	1.000	1.000	0.998	1.000	0.987	1.000	0.997
ST distance (10m buffer)	0.024	1.000	1.000	0.959	0.998	1.000	1.000	0.998	1.000	1.000	0.910	1.000
ST length (25m buffer)	0.06	1.000	1.000	0.910	1.000	0.982	1.000	1.000	1.000	0.998	0.781	1.000
ST distance (25m buffer)	0.110	1.000	1.000	1.000	1.000	0.740	0.985	0.999	1.000	0.953	1.000	0.997
ST length (35m buffer)	0.200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ST distance (35m buffer)	0.071	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.957	1.000
Ditch distance	0.066	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.988	0.849	0.999

Table 2. 15 ManyGLM analysis of area of patches for dominant vegetation and correlation with length of sheep track and distance to sheep track within a 10m, 25m and 35m buffer and distance to nearest ditch. ST = Sheep Track. Green = positive correlation, red = negative correlation. (p-values)

		Block formers						Opportunists				
Area	**overall**	**Callvulg**	**Eriovagi**	**Junceffu**	**Juncsqua**	**Molicaer**	**Nardstri**	**Carenigr**	**Descflex**	**Galisaxa**	**Poteerec**	**Vaccmyrt**
ST length (10m buffer)	0.430	1.000	0.979	0.996	1.000	0.892	1.000	1.000	1.000	0.966	1.000	0.998
ST distance (10m buffer)	0.085	1.000	1.000	0.944	0.998	1.000	1.000	0.962	1.000	1.000	0.998	1.000
ST length (25m buffer)	0.544	1.000	1.000	1.000	1.000	0.651	1.000	1.000	1.000	1.000	1.000	1.000
ST distance (25m buffer)	0.875	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ST length (35m buffer)	0.169	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ST distance (35m buffer)	0.023	1.000	1.000	1.000	1.000	1.000	1.000	0.961	1.000	1.000	0.866	0.996
Ditch distance	0.167	1.000	0.941	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.963	1.000

Table 2. 16 ManyGLM analysis of shape index of patches for dominant vegetation and correlation with length of sheep track and distance to sheep track within a 10m, 25m and 35m buffer and distance to nearest ditch. ST = Sheep Track. Green = positive correlation, red = negative correlation. (p-values)

		Block formers						Opportunists				
Shape	**overall**	**Callvulg**	**Eriovagi**	**Junceffu**	**Juncsqua**	**Molicaer**	**Nardstri**	**Carenigr**	**Descflex**	**Galisaxa**	**Poteerec**	**Vaccmyrt**
ST length (10m buffer)	0.533	1.000	0.976	1.000	1.000	0.976	0.624	0.998	1.000	1.000	1.000	1.000
ST distance (10m buffer)	0.478	0.999	1.000	1.000	1.000	0.579	0.997	0.999	1.000	0.992	1.000	0.999
ST length (25m buffer)	0.055	1.000	1.000	1.000	0.999	1.000	1.000	0.999	1.000	1.000	0.784	1.000
ST distance distance (25m buffer)	0.289	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ST length (35m buffer)	0.097	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ST distance (35m buffer)	0.059	0.979	1.000	1.000	1.000	1.000	0.972	1.000	1.000	1.000	0.996	1.000
Ditch distance	0.023	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.992

Table 2. 17 ManyGLM analysis of number of patches for subdominant vegetation and correlation with length of sheep track and distance to sheep track within a 10m, 25m and 35m buffer and distance to nearest ditch. ST = Sheep Track. Green = positive correlation, red = negative correlation. (p-values)

		Block formers						Opportunists				
No. of Patches	**overall**	**Callvulg**	**Eriovagi**	**Junceffu**	**Juncsqua**	**Molicaer**	**Nardstri**	**Carenigr**	**Descflex**	**Galisaxa**	**Poteerec**	**Vaccmyrt**
ST length (10m buffer)	0.436	0.955	0.390	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.997
ST distance (10m buffer)	0.048	0.630	1.000	1.000	1.000	1.000	1.000	0.930	1.000	1.000	1.000	1.000
ST length (25m buffer)	0.281	1.000	0.997	0.979	1.000	1.000	1.000	0.795	1.000	1.000	0.990	1.000
ST distance distance (25m buffer)	0.466	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ST length (35m buffer)	0.206	1.000	1.000	1.000	0.988	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ST distance (35m buffer)	0.205	0.986	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Ditch distance	0.063	1.000	1.000	0.910	1.000	1.000	1.000	1.000	1.000	0.997	0.882	1.000

Table 2. 18 ManyGLM analysis of area of patches for subdominant vegetation and correlation with length of sheep track and distance to sheep track within a 10m, 25m and 35m buffer and distance to nearest ditch. ST = Sheep Track. Green = positive correlation, red = negative correlation. (p-values)

		Block formers						Opportunists				
Area	**overall**	**Callvulg**	**Eriovagi**	**Junceffu**	**Juncsqua**	**Molicaer**	**Nardstri**	**Carenigr**	**Descflex**	**Galisaxa**	**Poteerec**	**Vaccmyrt**
ST length (10m buffer)	0.362	0.953	0.394	1.000	1.000	1.000	1.000	1.000	1.000	0.678	1.000	1.000
ST distance (10m buffer)	0.052	0.627	1.000	1.000	1.000	1.000	1.000	0.952	1.000	1.000	1.000	0.998
ST length (25m buffer)	0.346	1.000	1.000	0.984	1.000	1.000	1.000	0.762	1.000	1.000	0.966	1.000
ST distance distance (25m buffer)	0.400	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ST length (35m buffer)	0.137	1.000	1.000	1.000	0.995	1.000	1.000	1.000	1.000	1.000	1.000	0.997
ST distance (35m buffer)	0.196	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Ditch distance	0.108	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.937	1.000

Table 2. 19 ManyGLM analysis of shape index of patches for subdominant vegetation and correlation with length of sheep track and distance to sheep track within a 10m, 25m and 35m buffer and distance to nearest ditch. ST = Sheep Track. Green = positive correlation, red = negative correlation. (p-values).

		Block formers						Opportunists				
Shape	**overall**	**Callvulg**	**Eriovagi**	**Junceffu**	**Juncsqua**	**Molicaer**	**Nardstri**	**Carenigr**	**Descflex**	**Galisaxa**	**Poteerec**	**Vaccmyrt**
ST length (10m buffer)	0.362	0.953	0.394	1.000	1.000	1.000	0.981	1.000	1.000	0.995	1.000	1.000
ST distance (10m buffer)	0.052	0.627	1.000	1.000	1.000	1.000	1.000	0.952	1.000	1.000	1.000	0.998
ST length (25m buffer)	0.346	1.000	1.000	0.984	1.000	1.000	1.000	0.762	1.000	1.000	1.000	1.000
ST distance distance (25m buffer)	0.420	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.966	1.000
ST length (35m buffer)	0.085	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.984
ST distance (35m buffer)	0.202	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.984	1.000	1.000
Ditch distance	0.086	1.000	1.000	0.989	1.000	1.000	1.000	1.000	1.000	1.000	0.881	1.000

2.3.4 Interpolating patch metrics to the whole of Ashtrees Dipper (Objective 4)

A series of interpolation maps (Figs. 2.19 to Fig. 2.21) were generated for i) total number of patches, ii) mean area and iii) mean shape index of dominant and subdominant vegetation species. The total number of patches for dominant vegetation species ranges from about 16 to 36 while the total number of subdominant vegetation patches ranges from about 31 to 47. In general, interpolation maps show that a larger number of patches results in overall lower mean area. Note that in Figs. 2.19 to 2.21 the lower altitude area is at the top (North) of the maps (see also Fig 1.1 and Appendix 1.1 for soil pH, soil water, elevation and slope).

Dominant vegetation forms more patches at both intermediate and low altitudes (Fig. 2.19i) whereas patterns were less consistent for subdominant species (Fig. 2.19ii). For both dominant and subdominant species patch area was higher at high altitude (Fig. 2.20). Patch complexity had a similar range of values for both dominant and subdominant species (upper values of 1.52 and 1.41 respectively, Fig. 2.21) but there was considerably more variation across the heft in patch complexity, with some areas at the northern edge (lowest altitude) having a simple complexity (values near 1.0) with the rest of the heft showing a mixture of patches with both simple and complex shapes. Since dominant and subdominant species are often found within the same grid cell (10cm cells in 1m quadrat) there might be some overlap in the interpolation results, mostly notably number of patches (Fig. 2.19).

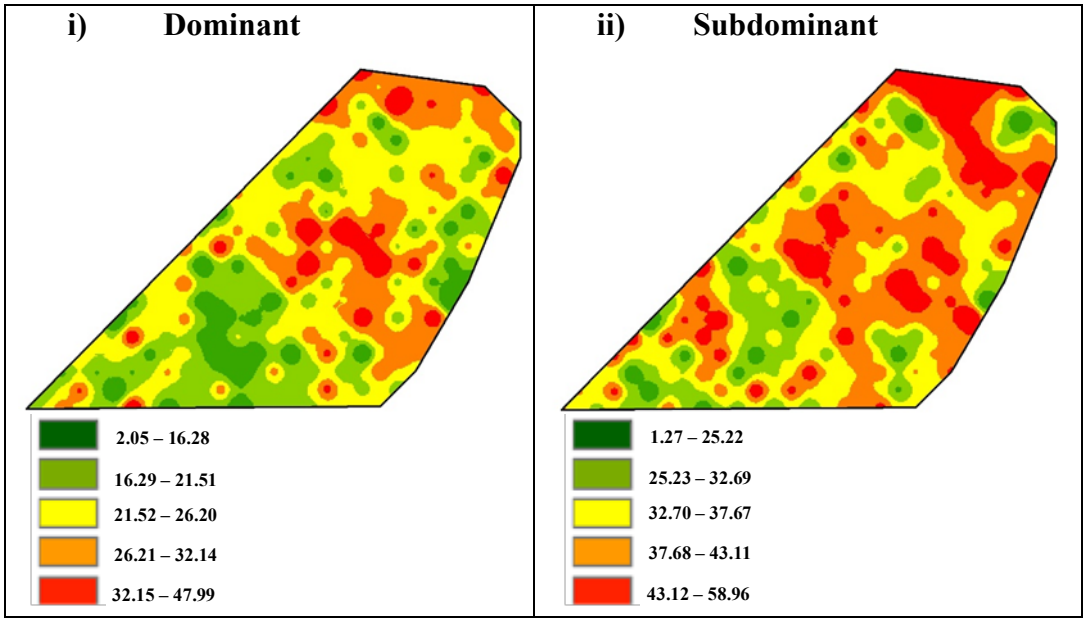


Fig. 2. 19 Interpolation of total number of patches for i) dominant (left) and ii) subdominant (right) vegetation across the Ashtrees heft for dominant and subdominant vegetation.

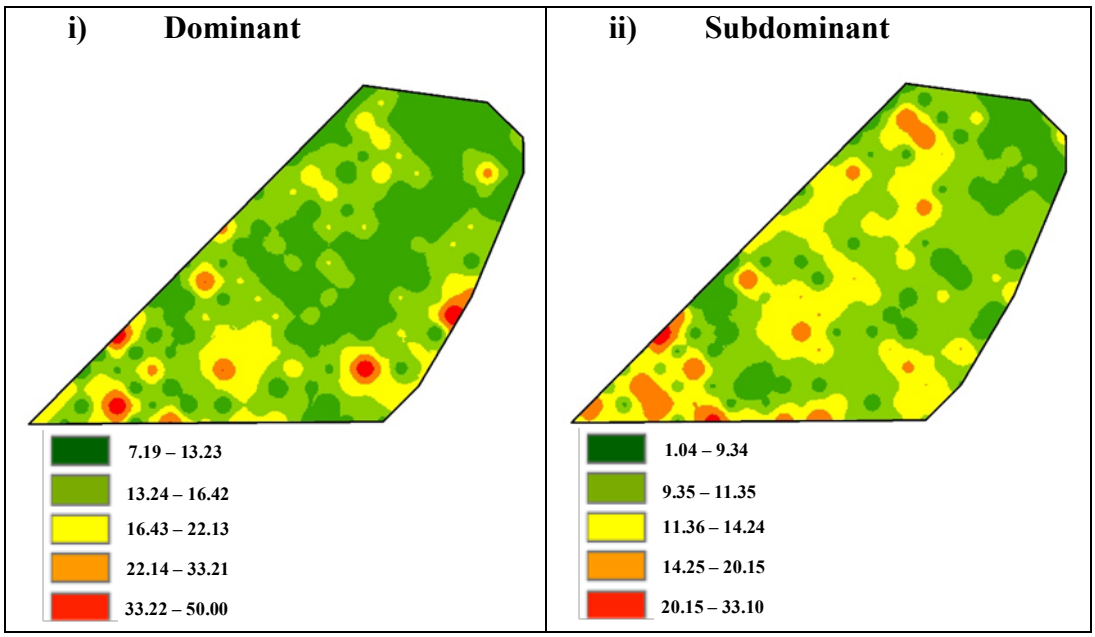


Fig. 2. 20 Interpolation of mean area of vegetation patches for i) dominant (left) and subdominant (right) vegetation across the Ashtrees heft for dominant and subdominant vegetation.

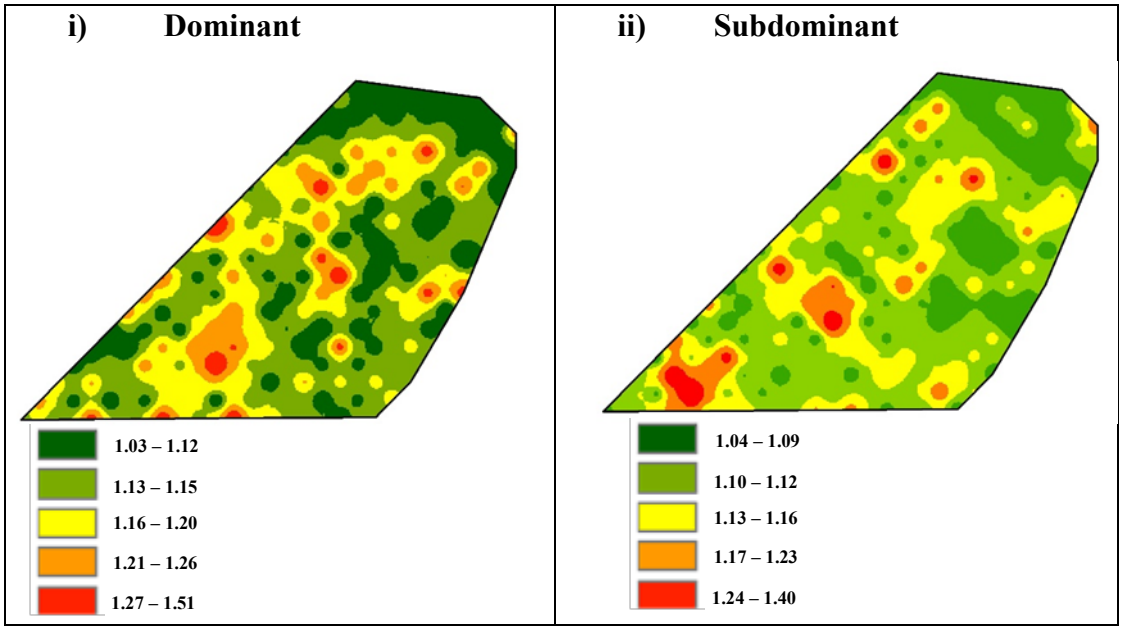


Fig. 2. 21 Interpolation of shape index data of patches for i) dominant (left) and ii) subdominant (right) vegetation across the Ashtrees heft for dominant and subdominant vegetation.

A series of interpolated maps for the block-former and opportunist species is provided in (Appendix 1.2 to 1.4). As an example, species such as *Juncus effusus* (Junceffu), *Juncus squarrosus* (Juncsqua) and *Nardus stricta* (Nardstri) form larger number of patches at lower altitudes. *Calluna vulgaris* (Callvulg) and *Carex nigra* (Carenigr) show more spread at higher altitudes. *Molinia caerulea* (Molicaer), *Deschampsia flexuosa* (Descflex), Galium saxatile (Galisaxa) *Vaccinium myrtillus* and *Eriophorum vaginatum* are more restricted in both distribution and patch metrics. Subdominant species within quadrats show similar distribution across the heft but show more patches but smaller in size.

Dominant species form simpler shapes than subdominant vegetation with the exception of a few species, e.g. *Molinia caerulea*. Subdominant opportunists, such as *Deschampsia flexuosa* and *Galium saxatile*, form very complex shapes and patterns across the heft. In general, the more complex the shape, the larger the area and but the lower number of patches formed.

2.4 Discussion

Landscape metrics have been used with the main aim of analysing species patches and their distribution at different scales in ecology (McGarigal, 2017). These metrics can be used to quantify composition, relationships and ecological processes both by comparing metrics and comparison with independent environmental data (McGarigal, 2017). This approach can be used to predict patterns across different spatial scales using interpolation techniques to gain an overall indication on how patches of species can vary across a field site in number, area and shape of vegetation patterns (Ritchie, 2009). It should be noted that the results obtained in this chapter, especially patch metrics at the species level, can be deemed as site-specific and might not reflect patch structure in other habitats. The methods and analysis in this chapter aim to understand and quantify vegetation patch structure at a $1m^2$ scale. The concept of 'micropatterning' of vegetation was introduced in the mid-1980s (Ohsawa, 1984) but the method did not compare vegetation patterns within different environments and differences between vegetation species (Ohsawa, 1984). Some studies, e.g. Berg *et al.* (1997) discussed how vegetation patchiness and fragmentation can be affected by external influences such as herbivore grazing. This, however, has not been quantified at very small spatial scales (Berg *et al.*, 1997).

Understanding how vegetation patches form requires integration of both ecological concepts such as management and environment with numerical analysis of the patch metrics (Kent *et al.*, 1997). Vegetation research has often focused on vegetation community composition which can respond rapidly to anthropogenic disturbance (Watt, 1947; van der Maarel, 1996). However, there have been fewer studies on the interactions between the individual species that collectively form the community and their environment (Bar Massada *et al.*, 2012).

2.4.1 Comparison between vegetation assessed via within-quadrat and survey methods (Objective 1)

Traditional vegetation surveys use percentage cover of all vegetation species to classify such vegetation into a community. This chapter shows that the use of dominant and subdominant data can provide a reliable understanding and indicative list of vegetation present and their respective community. NMDS results of the three data sets (percentage abundance, dominant and subdominant) show similarities in species distribution in ordination space along both axes

1 and axes 2. This similarity in R^2 results from Procrustes rotation and linear models show that surveys of dominant and subdominant vegetation species correlate well with abundance data and can still give reliable indication of vegetation community composition. This method was reported to also allow objective measure of the number of dominant and subdominant vegetation species since such species tend to be responsible for the majority of community make-up (Kikvidze and Ohsawa, 2002). The strong agreement between the dominant/subdominant approach to vegetation survey at 10cm scale with the traditional whole quadrat 1m scale increases the confidence in the methods used for the patch metrics, which rely on vegetation surveyed via the dominant/subdominant approach.

2.4.2 Relationship of dominant, subdominant and percent cover estimates at each quadrat with the environment (Objective 2)

Previous research (Sanderson *et al.*, 1995a) indicated that the differences in the actual numbers of livestock across the Ashtrees Dipper heft as a whole (in the early 1990's) did not have observable effects on the vegetation and thus sheep numbers were not used as a predictor. Furthermore, no data were available on the numbers of sheep active on for example the higher elevation areas of the heft compared to lower altitudes. Instead, sheep tracks were used as a surrogate for sheep grazing activity or the presence of sheep in an area. Soil water content and altitude significantly affected vegetation communities irrespective of whether cover estimates were derived from percentage abundance surveys, or dominant/subdominant data. The RDA plot for percentage cover shows correlation between altitude and soil water, but this is much weaker for the dominant/subdominant survey. This difference could be attributed to the fact that percentage cover data includes all species records (about 51 species) as opposed to filtering of the species in dominant (41 species) and subdominant (51) vegetation survey types. Both the percentage cover and dominant/subdominant surveys showed strong, but negatively correlated, effects from slope and soil water. This probably reflects flat areas tending to be more waterlogged, compared to faster water run-off on steep slopes (Klausmeier, 1999). This accords with other more general studies in which slope has been reported to have a strong effect on both the community composition and spatial pattern of vegetation patches formed (Watt, 1947; Coulson *et al.*, 1990*;* Klausmeier, 1999). The multivariate analyses suggest that soil pH also had strong effects on the community composition for both sampling techniques, which accords with a large body of previous literature (Miles, 1981; Goldberg, 1985; Sims, 1986; van Strien *et al.*, 1991; Dodd *et al.*, 1994; Pärtel *et al.*, 2004; Smith and Wyatt, 2007; Eskildsen *et al.*, 2013). However, soil pH

was not (inversely) related to soil water content as might have been expected. While soils at Ashtrees are relatively acidic, the situation has been made more complex by the inclusion of drainage ditches in some areas and attempts to improve the vegetation at lower elevations.

2.4.3 Environmental and management factors affecting the patch metrics of the quadrats (Objective 3)

2.4.3.1 Overall patch metrics

The number of patches formed significantly varied between dominant and subdominant vegetation types rather than between block-forming and opportunistic types. The lack of significance in the latter comparison could be because of the qualitative definition of these two broad vegetation types, and that a more precise definition of growth forms is required, possibly based on published vegetation traits. It was expected that block-forming species when a dominant cover type would produce fewer patches than opportunistic species when dominant. However, results show that both growth strategies produced similar mean numbers of patches in dominant species. The same is observed between block-formers and opportunists when species were subdominant. The major difference is between dominant and subdominant vegetation, where dominant species produce smaller number of patches than subdominant species, irrespective of growth strategy. However, whilst block-forming and opportunistic species can be considered as 'traits' applicable to individual species, the same is not true of the cover type, dominant versus subdominant. Some species were mainly recorded as dominants in a quadrat (e.g. *Hypnum cupressiforme*), others invariably recorded as sub-dominants (e.g. *Lolium perenne*, *Poa trivialis*) whilst some could be either dominant or subdominant, depending on their location within an individual quadrat (e.g. *Molinia caerulea*, *Carex nigra*). Nevertheless, those species that are primarily found dominant in a quadrat are more likely to be block-formers, whilst those primarily sub-dominant are more likely to be opportunists.

Dominant opportunistic species grow in significantly smaller patch areas than subdominant opportunistic species. This is because the opportunistic trait if observed in subdominant species forces them to grow within dominant vegetation as well as in any 'available' space within and around dominant vegetation (Addicott *et al.*, 1987; Grime, 1988). If an area of space is dominated by a block former of complex shape then one would expect any other (subdominant) block-formers in that area to produce relatively small sized patches.

Opportunists are adapted to colonising small areas, e.g. within the 'matrix' of the main block-former, or around the edges of block-formers (Klausmeier, 1999; van der Maarel, 1996; Bar Massada *et al.*, 2012). Therefore, in some situations, the areas of opportunists might actually be greater than that of subdominant block-formers. This type of process might underlie the results shown in Figs 2.17 and Fig. 2.18 in the results where patch area and shape were both significantly affected by cover type, growth strategy and the interaction between the two. For example, dominant block-formers might produce complex shapes (SI >>> 1; Fig. 2.22a) while subdominant block-formers produce comparatively simpler shapes (SI > 1; Fig. 2.22b) possibly because of lack of space and resources. Conversely, subdominant opportunists might grow in more complex shapes (SI >1; Fig. 2.22d) than dominant opportunists (SI ~ 1; Fig. 2.22c).

While this research set out to understand vegetation patch structure and any possible relationships with the primary environmental factors at Ashtrees, additional underlying features of the patch structure might arise from species phenology, functional traits, competitive abilities, regenerative strategies and seed dispersal. These species attributes have not been considered as they are out of scope of this research's aims. However, different plant species operate at different levels especially when taking into account the life history and the below-ground processes (Grime et al., 1988). For example, since upland environments can be exposed to harsh environmental conditions, regeneration capabilities (e.g. seedling, offspring persistence and growth), vegetation expansion and propensity to fragmentation of plants can be important factors that determine different levels of patchiness expressed by different species (Grime et al., 1988). These differences in expansion and fragmentation depend on both the characteristics of the individual plant species, and the environment in which it grows. For example, when conditions are favourable species form differing levels of root and rhizome networks that can very well be the main driver of patch composition and structure (Grime et al., 1988).

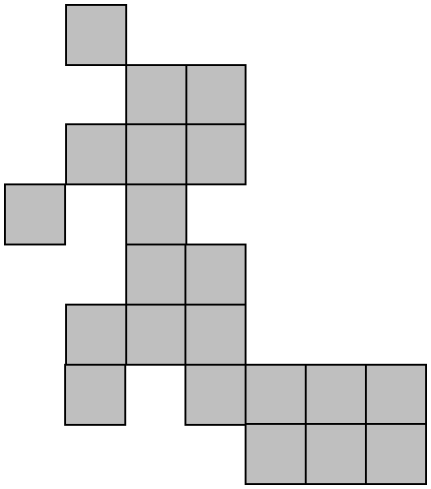
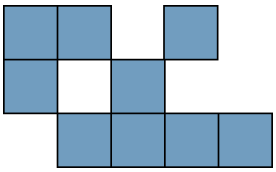
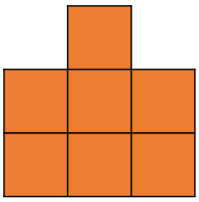
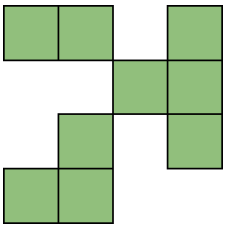
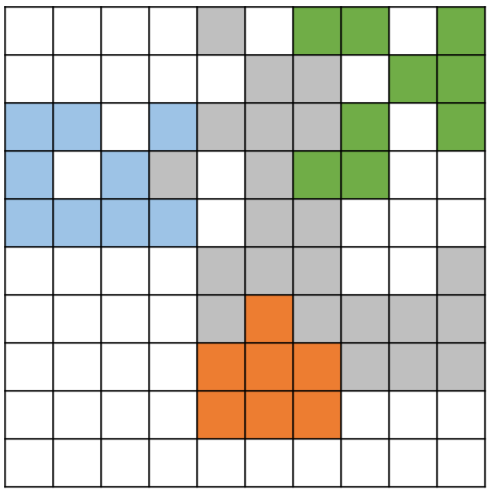
	Dominant	Subdominant
Block-formers	a) Large area, very complex shapes (SI>>>1) 	b) Medium area, complex shapes (SI>1) 
Opportunists	c) Small area, simple shapes (SI ~ 1) 	d) Medium area, complex shapes (SI>1) 
1m x 1m Quadrat; white areas occupied by other species		

Fig. 2. 22 Pictorial comparison of the difference of area and shape between different cover types and growth strategies of vegetation species (compare with Fig 2.17 and 2.18 in the Results).

2.4.3.2 Multivariate GLM analysis of patch metrics for all species in relation to the environment

Overall multivariate GLM analysis for dominant species showed that there were significant influences between individual species' number of patches and the environmental drivers. Patch area was correlated with soil pH and soil water content but not with slope. Subdominant vegetation species and their patch metrics showed higher correlation with environmental variables than dominant vegetation. It is interesting that while the number of patches formed by subdominant species is not statistically significant in their correlation to the altitude, the area and the shape of the patches are. This is probably because of limitations in resources and competition between the dominant species occupying the same area (Addicott *et al.*, 1987; Grime, 1988; Suding *et al.*, 2008).

The change in vegetation composition, structure, dynamics and pattern is evident in differences between lowlands and uplands, and these changes can be gradual or discontinuous over space (Ohsawa, 1984; Bruun *et al.*, 2006). While plant growth and development depends on numerous different environmental factors, altitude has major effects on number, area and shape of vegetation patches (Bruun *et al.*, 2006). For example, lowland grasslands, heathlands and moorlands can show substantial differences in structure and pattern when compared to those distributed in upland environments (Bruun *et al.*, 2006). In general, upland species grow in a seemingly narrower range of environmental conditions (such as soil pH, soil water etc.) than lowland species. However, vegetation in upland areas can adapt by shifting their distribution when environmental conditions change (Pateman and Hodgson, 2012). Research in the UK suggested that changes in climate and environmental resources, if exceeded a tolerance threshold, especially for grassland and heath vegetation, can substantially change in composition when temperature, rainfall and soil pH change (Ross et al., 2012). Many upland species are classified as 'stress-tolerators' (S) or intermediate between 'stress-tolerators' and 'competitors' (SC) under Grime, (1988) Competitor, Stress-Tolerator, Ruderal (CSR) system, indicating that competition does indeed occur between upland species. This competition may result in larger patches of more complex shapes in those species that dominate or outcompete others for space (Critchley *et al.*, 2002; Bruun *et al.*, 2006).

At Ashtrees Dipper soil pH becomes less acidic with increased altitude, which may explain why the soil pH and altitude show opposing trends for dominant/subdominant patch metrics. While soil pH might not be the sole driver of patch characteristics, the trends observed

indicate that there is an underlying influence by soil pH. This is especially true on areas of dominant species. Increased soil pH can decrease the patch area that species form but increases the number of fragmented patches across the Ashtrees field. This trend might be site-specific but has been reported to show similar trends in changes in patches at landscape-scale (Addicott *et al.*, 1987). The mean shape index of species when dominant does not seem to be affected by soil pH, even when the area for these species decreases (*pers. obs.*). Soil pH, unless exposed to major climatic, management or physical disturbance, does not alter over short time periods.

Altitude increases the importance of slope and aspect in determining vegetation growth and dynamic due to the different levels of solar radiation received by the plants (Bennie *et al.*, 2008). Slope, together with microclimatic pressures, influences plant dominance and distribution (Badano *et al.*, 2005). Topographic features can provide refugia and spatial variability that favour expansion of habitat range for certain species of plants (Badano *et al.*, 2005; Bennie *et al.*, 2006*;* Bennie *et al.*, 2008). The relationship between slope and vegetation dynamics is site-specific and therefore it can be difficult to predict the effects on patch metrics in different habitats (Bennie *et al.*, 2008). At Ashtrees the results indicated that subdominant patches are more affected by slope than dominant species (Table 2.8 to 2.13, Figs. 2.14 to 2.15). Furthermore, both patch shape became less complex, and overall patch area decreased with increased slope, which suggests that both dominant and subdominant species become more fragmented with increase in slope (Table 2.8 and Table 2.11). The underlying process behind this effect is unclear.

Surface water runoff and groundwater are affected by meteorology, topology and soil type (Wigmosta *et al.*, 2002). Different species will be better adapted to wet or dry soils, which may, in turn, affect which ones become dominant or subdominant in a given location because of interspecific competition. In areas with both high soil water content and high altitude vegetation patches become larger and more complex (Klausmeier, 1999). The results show that at the Ashtrees Dipper, soil water content has a positive correlation with all overall patch metrics and shows some correlation with specific vegetation species, namely *Eriophorum vaginatum* (from the selected species in this research). This effect might be driven by other species present at the study site that were not selected as part of this study.

2.4.3.3 Sheep tracks and ditches

The number of patches formed by dominant/subdominant vegetation species increased when sheep tracks occurred within 10m of the quadrat. In addition, species when dominant had increased shape complexity when in close proximity to sheep tracks. These two findings suggest that high grazing pressure on both dominant and subdominant species increases vegetation fragmentation (Thomas, 1959; Plumptre, 1994; Maron and Crone, 2006; Stewart and Pullin, 2006). Species near sheep tracks might be less palatable to sheep and more resistant to trampling, which might also increase shape complexity of the nearby vegetation (Maron and Crone, 2006). Furthermore, increases in pattern complexity has been reported as an 'evasion' strategy to optimise chances of survival for some species in areas of high grazing effort (Fisher *et al.*, 1996). This 'evasion' has been observed in some of the more palatable species (e.g. *Caluna vulgaris*, *Nardus stricta*) allowing other species such as *Agrostis capillaris*, *Deschampsia flexuosa*. *Festuca rubra* etc. to grow in proximity to sheep tracks while the palatable species grow further away from the tracks (Fenton, 1937; Krahulec *et al.*, 2001)

Drainage ditches are no longer actively dug across large areas of the UK uplands but extant drainage networks still affect vegetation growth (Holden *et al.*, 2004; Ramchunder *et al.*, 2009). However, there were no consistent associations between proximity to drainage ditches and patch metrics at Ashtrees. The presence of ditches reduces the abundance of mosses and unpalatable grasses such as *Nardus stricta* (Coulson *et al.*, 1990; Ramchunder *et al.*, 2009). The overall number of patches formed by dominating species decreased when in close proximity to a ditch (Table 2.14 and Table 2.17); the mechanisms behind this are unclear, but might arise from lower soil water favouring other species.

2.4.4 Interpolating patch metrics to the whole of Ashtrees Dipper (Objective 4)

Subdominant vegetation produce more patches than dominant vegetation species (Fig. 2.19 and Appendix 1.2). This could be as result of increased fragmentation of subdominant vegetation in areas of the heft where dominant vegetation produces large patches. Subdominant vegetation will grow in small spaces within or around dominant vegetation patches to maximise space and resource utilisation (Suding *et al.*, 2008). Species at Ashtrees such as *Juncus* species, *Potentilla erecta* and *Vaccinium myrtillus* that were less frequently classified as dominants or subdominants tended to form very small number of patches. For

example, *J. effusus* and *J. squarrosus*, species of different ecological requirements, occurred together in the northern part of the field (Appendix 1.2) and this occurrence could possibly be associated with small streams (pers. obs.). When species are subdominant they were more widely distributed across Ashtrees Dipper, and formed slightly higher number of patches. Conversely, when species are dominant they form larger and more complex areas (Section 2.4.3.2).

2.5 Conclusion

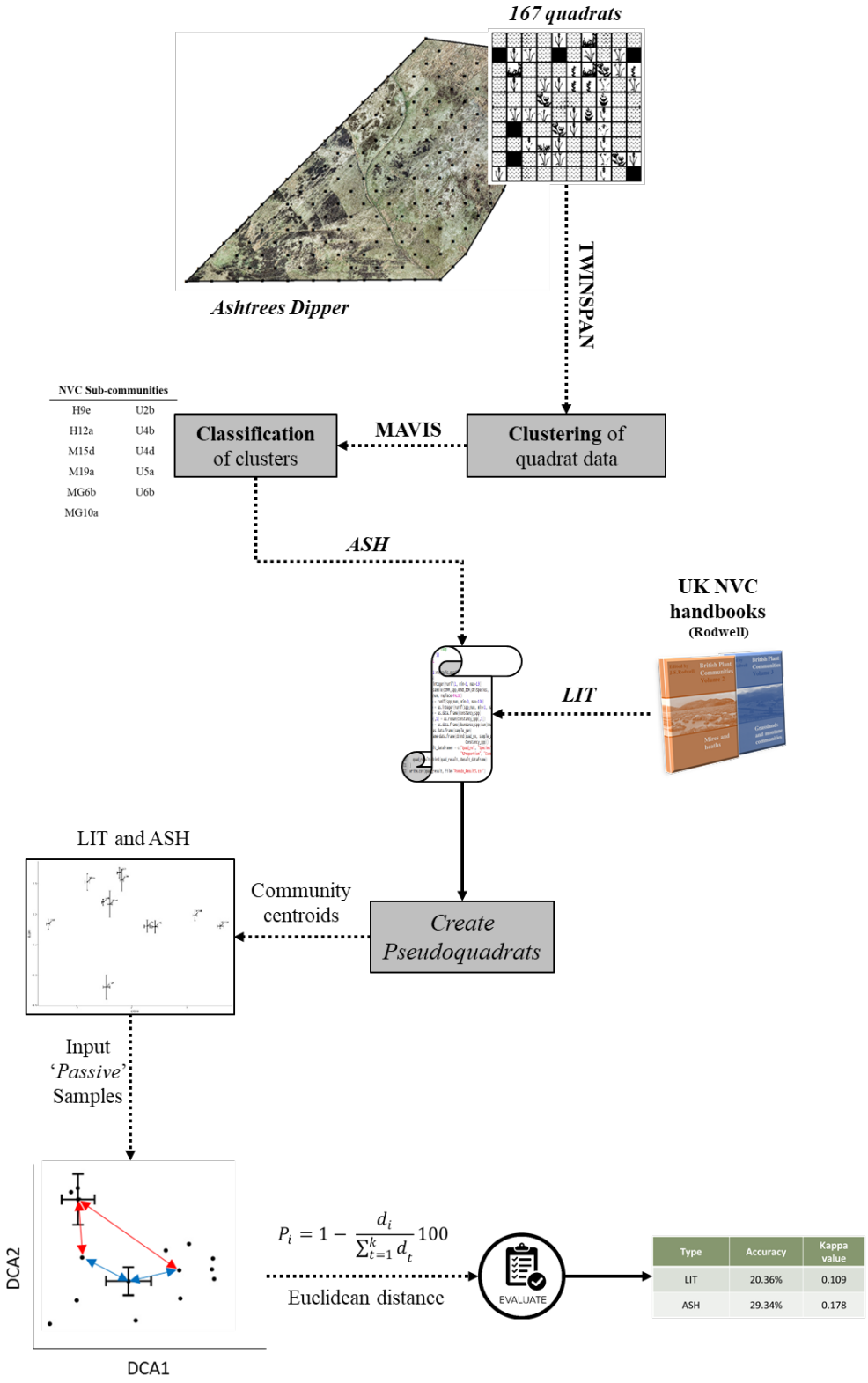
This research indicates that dominant and subdominant vegetation sub-quadrat data is comparable with that obtained via traditional survey methods of percentage abundance when scaled to the whole-quadrat. Whilst the dominant/subdominant method does not record rarer species within a quadrat, it nevertheless provides a robust survey technique that also provides insights into vegetation patch structure.

RDA and multivariate GLM results show environmental conditions (soil pH, soil water content, slope and altitude) have a significant effect on vegetation, irrespective of their cover type (dominant/subdominant) or growth strategy (block-formers/opportunists). Comparison between vegetation cover type and growth strategy resulted in a significance in both patches area and shape formed by block-forming and opportunistic dominant and subdominant vegetation species. Sheep paths show a more limited influence on vegetation patch structure, affecting number, area and shape of patches of dominating species more than subdominant vegetation. The presence of drainage ditches showed no substantial influence on vegetation of either cover type, only affecting number of patches by dominant vegetation, and certain individual species. Patch metrics of dominant and subdominant vegetation, and associated environmental data, can be interpolated from the 10cm or quadrat scale to the whole field to aid interpretation.

Chapter 3. Generalisable methods to classify vegetation using 'pseudoquadrats'[1]

[1] Accepted and presented as a peer-reviewed conference paper at the 9th International Congress on Environmental Modelling and Software at Colorado, USA, June 2018.

Graphical Abstract



Abstract

Many countries have developed phytosociological classifications of their vegetation. Can methods be developed for any classification system to allocate newly surveyed quadrats into the most likely vegetation community? Algorithms or software already exist to allocate quadrats for some national classifications, but these are not generalisable to any system. This chapter tests the robustness of generalisable approaches to allocate quadrats to an existing phytosociological classification, using the British National Vegetation Classification (NVC) as a case study.

Vegetation from 167 quadrats from a survey at Ashtrees Dipper farm in Northumberland, UK, was used for this case study. The vegetation quadrats were classified using two-way indicator species analysis (TWINSPAN) and the resultant groups allocated to communities within the NVC using the NVC-specific 'MAVIS' software. These were considered the most accurate description of the observed communities. Sets of 25 artificial 'pseudoquadrats' for each of these communities were computer-generated based on either the published lists of species in the relevant NVC community descriptions or from the list of species actually surveyed at the Ashtrees Dipper site. Distance in ordination space of observed quadrats from pseudoquadrats was used to predict community type.

Eleven sub-communities were observed at the site identified via TWINSPAN+MAVIS. Confusion matrices to compare predicted vs observed gave 20% accuracy for pseudoquadrats created from the literature, and 30% for pseudoquadrats generated from the subset of species known to occur at the site. While the results of accuracy are still fairly low, overall results demonstrate that the use of pseudoquadrats provides a flexible, generalisable means to objectively allocate vegetation quadrats into any extant classification system.

3.1 Introduction

3.1.1 Phytosociology and community classification

Phytosociological vegetation classification systems have been developed in numerous countries to describe semi-natural and natural vegetation communities or relevés. These include the regional European Vegetation Archive (EVA) currently being developed for Europe and neighbouring countries (EVS-European Vegetation Survey, 2018), the Irish Vegetation Classification (IVC: Biodiversity Ireland, 2018), the United States National Vegetation Classification (USNVC, 2016), the Canadian National Vegetation Classification (CNVC; Canadian National Vegetation Classification, 2013), the New Zealand National Vegetation Survey Databank (NVS; Landcare Research, 2016), and the Great Britain National Vegetation Classification (NVC – Rodwell, 1998a; Rodwell, 1998b; Rodwell, 1998c; Rodwell, 2006). There are broad similarities in the structure of some of these systems, for example hierarchical classes of different vegetation reléves into 'communities' and 'sub-communities' (British NVC) or broader 8-level 'formations' through to 'associations' (USNVC; CNVC). Some countries have collaborated to produce standard methods for vegetation database management, e.g. TurboVeg recommended by the EVS and International Association for Vegetation Science (Hennekens and Schaminée, 2001).

Due to the different range of habitats and ecosystems encompassed by these vegetation classifications, they differ in both their recommended field survey methods, and in the techniques used to allocate quadrats to vegetation classes. For example, the British NVC recommends a minimum of five quadrats per relevé, precluding the use of many historical datasets where vegetation was surveyed on a per-quadrat level, whereas the Irish IVC can use single quadrats. Some classifications do not provide formal methods to allocate new quadrats to classes, whilst software has been developed for some national systems, e.g. ERICA for the IVC. The British NVC is derived from a two-way indicator species analysis (TWINSPAN, (Hill, 1979), and the resultant published NVC handbooks provide paper-based keys, analogous to binomial taxonomic keys, to allocate quadrats to communities. To ease classification computer-based methods to allocate field quadrats have been developed including MATCH (Malloch, 1998), TABLEFIT (Hill, 1989; Dodd *et al.*, 1994; CEH, 2014) and most recently the Modular Analysis of Vegetation Information System (MAVIS; Smart *et al.*, 2016).

Whilst such computer-based systems have been shown to be invaluable aids for vegetation scientists, several problems remain. First, their design may preclude the allocation of some historical vegetation records within a national classification due to differences in survey methods, for example the need for a recommended minimum of five random quadrats per relevé for the British NVC. Quadrats may have been placed at regular intervals along transects, across survey grids, or at random, none of which accords with standard NVC survey methods. Subsequently ecologists, especially when utilising published data from historical surveys, may wish to incorporate such quadrats into a national system and therefore need to be able to resolve differences in survey methods. Second, such computer-based systems are not available for many national systems, and even where software is available, outputs are usually restricted to tables that indicate the highest predicted vegetation classes. In the real world, semi-natural and natural vegetation generally exists in a continuum, therefore it would be useful to be able visualise this continuum in an ordination diagram that can be readily interpreted by practicing vegetation scientists, whilst still obtaining the advantages of a national classification system. Such a system has to be relatively simple, so that it could be generalised to any country's vegetation classification.

3.1.2 The UK National Vegetation Classification (NVC)

Vegetation classifications should provide a means to identify similarities (or differences) between vegetation species that are spatially separated on both small- and large-scales (Dodd *et al.*, 1994). The UK NVC is the mostly widely used phytosociological classification system in the UK (Hearn *et al.*, 2011). The NVC's comprehensive analysis of vegetation communities and their species composition aids interpretation of relationships between climatic, physical and biotic factors in different vegetation-dominated environments (Rodwell, 2006). The NVC communities and their distribution were published in a series of 'British Plant Communities' books (Rodwell, 1998a; Rodwell, 1998b, Rodwell, 1998c). Scientists have used this extensive floristic database in different analytical and/or computational techniques for environmental conservation and management (Morecroft *et al.*, 2009; Bradter *et al.*, 2011). This chapter uses vegetation subcommunity data from mires and heaths (volume 2; Rodwell, 1998b) and grasslands and montane communities (volume 3; Rodwell, 1998c).

3.1.3 Modular Analysis of Vegetation Information Systems (MAVIS)

The most modern computerised vegetation classification for the UK NVC is 'Modular Analysis of Vegetation Information System (MAVIS), developed by DART Computing and designed by Simon Smart at the Centre for Ecology & Hydrology (CEH; Smart *et al.*, 2016). MAVIS is a successor to MATCH (Malloch, 1998) and TABLEFIT (Hill, 1993). The software provides an objective classification of vegetation data through the matching of vegetation communities classified by the NVC, according to either abundance or constancy values (Dodd *et al.*, 1994; Smart *et al.*, 2016). A weakness of precursors to MAVIS is that they were dependent on the availability of constancy values (I-V) of each species in the data sample (Dodd *et al.*, 1994).

MAVIS is based on the Countryside Vegetation System (CVS; Bunce *et al.*, 1999) and the National Vegetation Classification (NVC; Smart *et al.*, 2016). The Countryside Vegetation System is based on a series of surveys to provide a broad classification of vegetation environments, that divides the British landscape into a number of land classes that also include hedgerows, stream-sides and road verges (Blain, 2009; Smart *et al.*, 2016). Unlike MATCH, MAVIS allows input of data of species lists with either constancy or abundance values as opposed to solely constancy values (Smart *et al.*, 2016). Even though constancy values were previously used in MATCH as the main input to differentiate plant communities, the use of % abundance of each species allows a more accurate and realistic description of vegetation community composition (Dodd *et al.*,1994). The resultant outputs (from MAVIS) provide matching coefficients between the input data and the communities defined in the National Vegetation Community (Dodd *et al.*, 1994, Smart *et al.*, 2016). In addition, MAVIS also outputs results of (i) Ellenberg and (ii) Competitor, Stress-tolerators and Ruderal scores associated with input species (Smart *et al.*, 2016) as well as Ellenberg scores based on the proportion of each species found within a community (Morecroft *et al.*, 2009, Smart *et al.*, 2016).

Abiotic factors such as soil pH, soil fertility, wetness and light requirements and altitude (Malik and Husain, 2006), as well as interactions with neighbouring plant species (Rodwell, 1998a; Rodwell, 1998b, Rodwell, 1998c; Smart *et al.*, 2016) all affect the observed community composition.

3.1.4 Phytosociological Ordination techniques

Ordination techniques have been widely used to simplify complex multivariate data, such as samples by species tables, into a low number of ordination axes that summarise the main patterns in the data. There has recently been development of new multi-species analyses, such as model-based methods (Warton *et al.*, 2015) to include different error distributions (e.g. binomial for presence/absence, Poisson for count data, and Gaussian for continuous data), or Bayesian methods to account for imperfect detection in surveys (Warton *et al.*, 2015; Beissinger *et al.*, 2016; Warton *et al.*, 2016). However, some of these newer methods are more expensive to implement, for example imperfect detection techniques require multiple site visits and are probably less essential for sessile organisms such as vegetation (Warton *et al.*, 2016). Most vegetation ecologists therefore still mainly use standard 'algorithmic' ordination techniques, such as principal components analysis (PCA), correspondence analysis (CA), detrended correspondence analysis (DCA) and non-metric multidimensional scaling (NMDS). None of these methods can be used for all multivariate datasets, as their underlying methods and assumptions differ. For example, PCA assumes a linear relationship between the derived axes and the abundance of the species, whereas CA and DCA assume unimodal relationships (Ter Braak, 1987), which are sometime more suited for 'sparse' species datasets where a few common species predominate, and there are many rare species. In general, the choice between using linear (e.g. PCA) or unimodal (CA/DCA) ordination techniques can be based on the length of the gradient seen in the data (i.e. length of the first ordination axis; see Lepš and Šmilauer, 2003; Jongman et al., 1987 and Section 1.3). This choice of ordination can also depend on the type of data and the variation within the data. In extreme cases, where the raw data is dominanted by a few outlier species, some ordination methods do not provide meaningful visualisation, or transformations of the raw data are required before analysis (see Legendre and Legendre, 2014).

One problem that can arise with both PCA and CA is the so-called 'horseshoe effect' where samples relatively dissimilar in their species composition can be placed relatively close together in ordination space, due to an artefact causing samples to be displayed in an arch. DCA tries to correct this artefact by splitting the first ordination axis into segments, and detrending the samples within each segment to remove the arch. NMDS is computed from a similarity matrix of the original species data, typically the Bray-Curtis similarity index, and re-projects samples and species onto a pre-defined number of axes (usually two). Irrespective of the ordination method used, new samples, sometimes referred to as 'passive samples', that were not included in the original ordination, can be placed into the extant ordination space

without altering the original configuration of samples (Sanderson and Rushton, 1995). Ordination is often used in conjunction with classification techniques to group together related samples (and species) to aid interpretation. DCA and NMDS provided the most visually promising ordination plots for the vegetation data described in this chapter, and two-way indicator species analysis (TWINSPAN – Hill 1979) was used as the classification technique, as this was used in the original development of the NVC.

3.1.5 Aims and objectives

The overall aim of this chapter was to test the robustness of several generalisable approaches to allocate quadrats to an existing phytosociological classification, using the British National Vegetation Classification (NVC) as a case study.

The specific objectives were to:
1) produce an initial baseline NVC classification of individual quadrats at Ashtrees Dipper. This uses multivariate TWINSPAN classification followed by MAVIS to provide the most reliable "observed" allocation of these quadrats to an NVC community (Objective 1)
2) develop a method to generate artificial vegetation quadrats ('pseudoquadrats') for subsequent analysis. These pseudoquadrats are based on the species lists for each community identified in Objective 1, for either the species in the NVC handbooks ('literature data') or a more restricted species list based on those observed at Ashtrees. Both sets of pseudoquadrats can then be analysed via conventional ordination methods to provide a visual display of the NVCs that occur at the site. (Objective 2)
3) allocate field quadrat data as passive samples into these ordinations and calculate predicted NVC membership and compare these with "observed" NVC memberships identified in Objective 1.
4) use interpolation to predict communities at across the whole of Ashtrees Dipper heft (Objective 4)

3.2 Methodology

3.2.1 Produce an initial "baseline" for NVC classification (Objective 1)

Initial classification of the vegetation was done using MAVIS using abundance data gathered from Ashtrees in 1991 (167 1m^2 quadrats). This was done to try and produce the most accurate 'baseline' classification of the NVC communities for all the quadrats at the site against which to compare the new pseudoquadrat techniques, since MAVIS is the current classification software used in the UK (Smart *et al.*, 2016). However, MAVIS requires a minimum of 5 quadrats to produce the most accurate classifications (ideally 4m^2 quadrats) rather than single 1m^2 quadrats. Therefore, an initial classification of all 167 quadrats was performed using Two-way Indicator Species Analysis (TWINSPAN) hierarchical classification technique (Hill, 1979; Roleček *et al.*, 2009). TWINSPAN was chosen as this is the classification method used during the original development of the British NVC. TWINSPAN allocated the 167 quadrats to 8 clusters using Whittaker dissimilarity index. Whittaker dissimilarity index provided the best clustering, with no one quadrat having less that 5 quadrats in each cluster. The quadrats in these 8 clusters were then inputted into MAVIS and the predicted communities and sub-communities derived. In cases when MAVIS classified quadrats into different communities but with identical matching coefficients both communities were allocated. This resulted in eleven sub-communities identified to occur at Ashtrees Dipper. These eleven sub-communities were then used as the subset to create and test the generalisable pseudoquadrat methods.

3.2.2 Generate 'pseudoquadrats' from literature and survey data (Objective 2)

The method for developing pseudoquadrat techniques was originally described in Sanderson *et al.* (1995a). The new methods developed in this research provide a more readily generalisable to classify vegetation quadrats within any country's classification system. Two complementary approaches were used to generate the pseudoquadrats for a sub-community. The first method was based on the entire set of species published in the UK NVC handbooks for a sub-community (henceforth denoted LIT), whilst the second was restricted to utilise only the set of species recorded in the Ashtrees field survey (ASH).

Twenty-five pseudoquadrats were generated for each sub-community (for both LIT and ASH), with the number of species within a pseudoquadrat being determined by two different

randomisation methods. In the LIT method the number of species within each pseudoquadrat was derived by randomly selecting a number between the published minimum and maximum expected number of species per quadrat. For the ASH approach this was based on the minimum and maximum number of species per quadrat recorded across all the field quadrats at Ashtrees Dipper.

In the LIT method the species were randomly selected from all the species recorded within that sub-community. . The species percentage cover abundance reported in the NVC handbooks for that sub-community was then allocated to the species. In the ASH method, the species were randomly selected from the list of species surveyed at Ashtrees. The randomly selected species was allocated the same percentage cover abundance value that it had in the survey. The random selection (without replacement) of species was repeated until the total number of species in a pseudoquadrat reached that initially drawn. This randomisation process was performed for both the LIT and ASH methods. The 275 pseudoquadrats (11 sub-communities, 25 pseudoquadrats each) were ordinated by detrended correspondence analysis (DCA – Hill, 1979) since DCA provided a clearer distribution of pseudoquadrats in ordination space. The mean centroids and standard errors of each sub-community for both the LIT and ASH quadrats were visualised in ordination space.

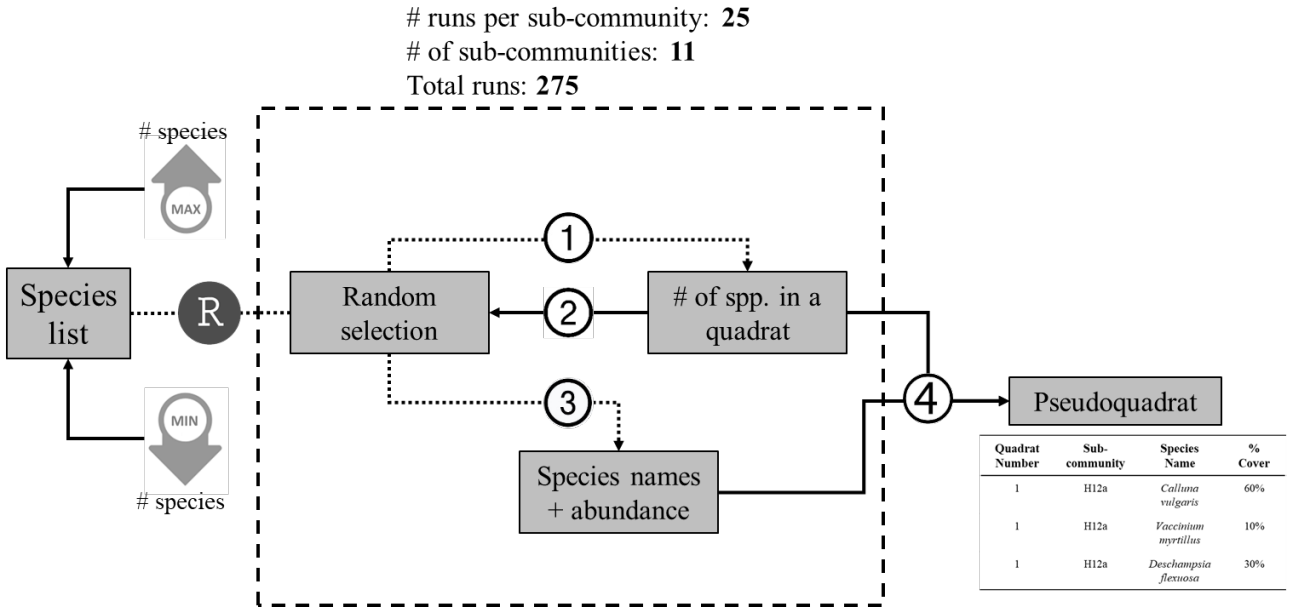


Fig. 3. 1 Graphical summary of the randomisation procedure to generate LIT and ASH pseudoquadrats.

3.2.3 Allocate and predict field quadrat data to ordinations (Objective 3)

Field quadrats were placed within the both the pseudoquadrat DCA LIT and DCA ASH ordination spaces as 'passive samples'. The positions of the passive samples were based on the pseudoquadrat ordination scores, and passive samples (field quadrats) had no effect on the original ordination (Hill, 1979; Sanderson *et al.*, 1995b). The positions of passive samples are calculated from the species scores of the 'real' quadrat data; this ensures that the scores from the 'real' quadrat ordination can be re-projected into the original ordination space that also included the 'pseudoquadrat' centroids (see Jongmann et al. 1991). The distance of the field quadrats from each sub-community mean centroid was calculated. It was assumed that the shorter the Euclidean distance between a field quadrat and a mean centroid, the higher the probability that the field quadrat belonged to that sub-community. The following equation, modified from that in Sanderson *et al.*, (1995a), was used to determine the probability that a field quadrat belonged to a sub-community:

$$p_i = 1 - \frac{di}{\sum_{t=1}^{k} dt}\ 100$$

where,

p_i = probability that field quadrat belongs to community i

d = distance of field quadrat to mean centroid of community

k = total number of vegetation communities

All analyses were undertaken in R (Version 3.4.2, Vienna, Austria) plus the vegan (Oksanen, 2015) and TwinspanR packages (Roleček *et al.*, 2009). To assess the reliability of the methods developed as well as the overall reliability of MAVIS, two confusion matrices were produced:

i) MAVIS vs LIT

ii) MAVIS vs. ASH

Confusion matrices are tables that are used to calculate measurements of performance of new or developed techniques by comparing observed and predicted values (Narkhede, 2018). These tables operate by comparing 'agreement' between rows and columns. In this case each column is a community classification produced by MAVIS while each row is a community classification obtained from each of LIT or ASH ordination methods. Cohen's Kappa values were used as a measure of accuracy of the LIT/ASH method with the 'baseline' classification.

3.2.3 Map community distributions across Ashtrees Dipper (Objective 4)

The classification of each quadrat using the ASH method was then used to interpolate the vegetation communities across the Ashtrees field. Inverse distance weighting (IDW) interpolation method was used and performed using ArcGIS (ESRI, 2015). A series of 11 interpolation maps was obtained, one for each identified vegetation community.

3.3 Results

3.3.1 Produce an initial "baseline" for NVC classification (Objective 1)

TWINSPAN hierarchical classification resulted in 8 clusters (Fig. 3.2). Classification of these 8 clusters by MAVIS resulted in 11 possible sub-communities. This is because the classification of three TWINSPAN clusters resulted in an additional 3 sub-communities in which scores were identical. The eleven sub-communities derived from TWINSPAN + MAVIS are summarised in Table 3.1.

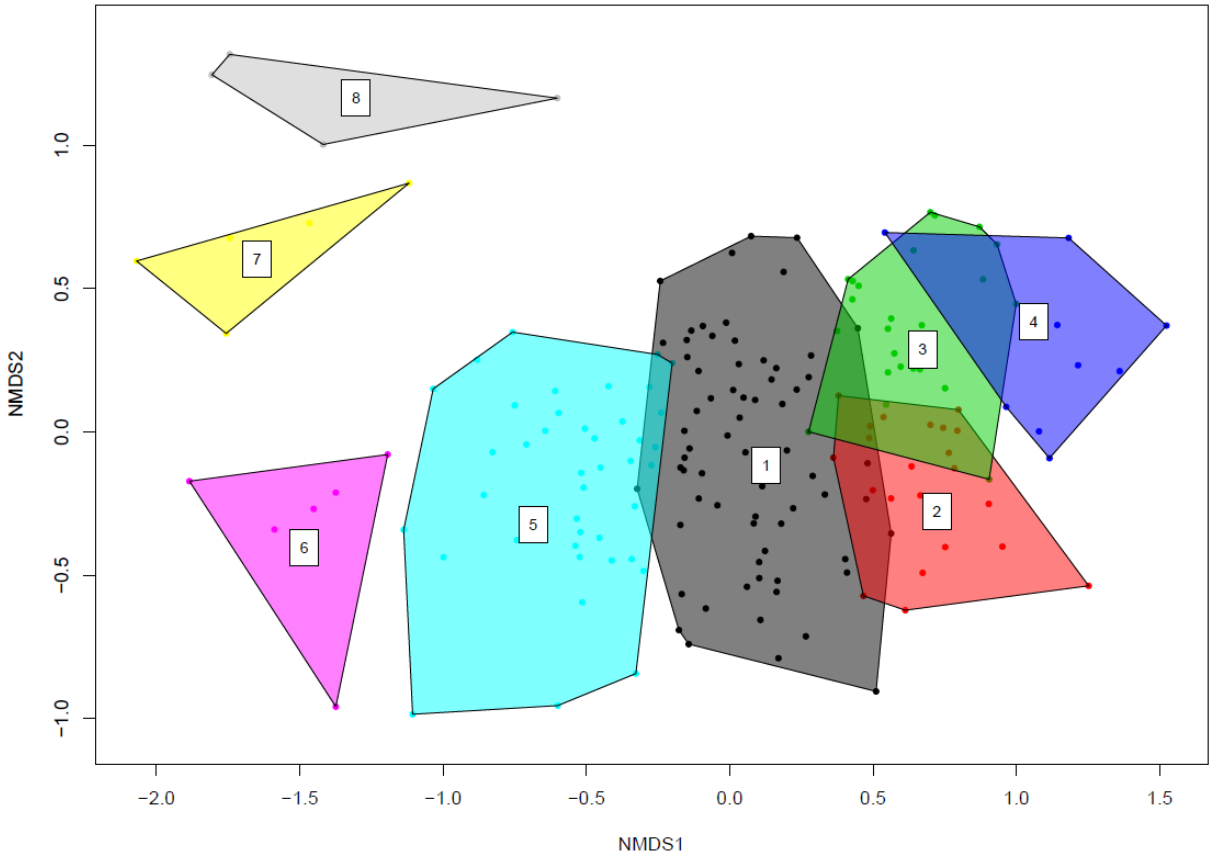


Fig. 3. 2 Hierarchical cluster plot for NMDS analysis. Clusters were produced using TWINSPAN hierarchical analysis in R (R Core Team, 2013)

.

Table 3. 1 TWINSPAN cluster number, community and sub-community codes and name of each community and sub-community and the number of quadrats per sub-community found at Ashtrees.

Cluster	Community code	Community type	Sub-community code	Sub-community type	Matching Coefficient by MAVIS
1	U2	*Deschampsia flexuosa* grassland	U2b	*Vaccinium myrtillus* sub-community	60.03%
1	U6	*Juncus squarrosus-Festuca ovina* grassland	U6b	*Carex nigra – Calypogeia trichomanis* sub-community	60.03%
2	M15	*Scripus cespitosus-Erica tetralix* wet heath	M15d	*Vaccinium myrtillus* sub-community	51.25%
3	H12	*Calluna vulgaris-Vaccinium myrtillus* heath	H12a	*Calluna vulgaris* sub-community	51.84%
4	M19	*Calluna vulgaris-Eriophorum vaginatum* blanket mire	M19a	*Erica tetralix* sub-community	54.87%
5	U5	*Nardus stricta-Galium saxatile* grassland	U5a	Species-poor sub-community	59.36%
6	U4	*Festuca ovina-Agrostis capillaris-Galium saxatile* grassland	U4b	*Holcus lanatus – trifolium repens* sub-community	47.96%
6			U4d	*Luzula multiflora – Rhytidiadelphus loreus* sub-community	47.96%
7	MG10	*Holcus lanatus-Juncus effusus* rush pasture	MG10a	Typical sub-community	47.34%
7	MG6	*Lolium perenne-Cynosurus cristatus* grassland	MG6b	*Anthoxanthum odoratum* sub-community	47.34%
8	H9	*Calluna vulgaris-Deschampsia flexuosa* heath	H9e	*Molinia caerulea* sub-community	40.00%

H: Heaths

M: Mires

MG: Mesotrophic grasslands

U: Calcifugous grasslands/montane communities

Rodwell, J. S. (1998a). British Plant Communities Volume 2 Mires and heaths. Cambridge, UK, Cambridge University Press.

Rodwell, J. S. (1998b). British Plant Communities Volume 3 Grassland and montane communities. Cambridge, UK, Cambridge University Press.

The highest matching cluster by MAVIS was cluster 1 matching at 60%, classifying this set of quadrats as U2b and U6b sub-community. Cluster 5 was also a fairly well-defined sub-community (U5a), scoring 59.36% matching with NVC data. This was followed by cluster 4, classifying the quadrat set within M19a sub-community. The remaining clusters resulted in

low matching scores with NVC data within the MAVIS database. Clusters 6 and 7 resulted in two possible sub-communities. Cluster 6 was classified as either U4b or U4d sub-communities while cluster 7 either a MG10a or MG6b sub-communities. The H9e sub-community (cluster 8) was the least well-defined (matching coefficient = 40.00%).

3.3.2 Generate pseudoquadrats from literature and surveyed data (Objective 2)

Centroids of sub-communities for pseudoquadrats derived from the LIT and ASH methods are summarised in Figs. 3.3 and 3.4 respectively. In both ordinations DCA axis 1 represents a trend from the higher altitude, acid sub-communities (H and U) to mesotrophic grasslands (MG6b and MG10a). DCA axis 2 separated H12a *Calluna vulgaris-Vaccinium myrtilis* heath from the remaining acid grassland and heaths in the LIT method. H12a is totally dominated by *Calluna vulgaris* which can represent over 90% of the vegetation cover (Rodwell, 1998a). H12a and M15d *Scirpus cespitosus-Erica tetralix* wet heath produced the most variable pseudoquadrats, as represented by their error bars especially on axis 2 for the LIT method (Fig. 3.3). There was little variability amongst the pseudoquadrats generated by the LIT method for U6b *Juncus squarrosus-Festuca ovina* grassland. In the ASH method (Fig. 3.4) axis 2 separates H9e from the rest of the sub-communities. H9e is the *Molinia caerulea* sub-community within H9 *Calluna- vulgaris-Deschampsia flexuosa* heath, and is relatively species-poor.

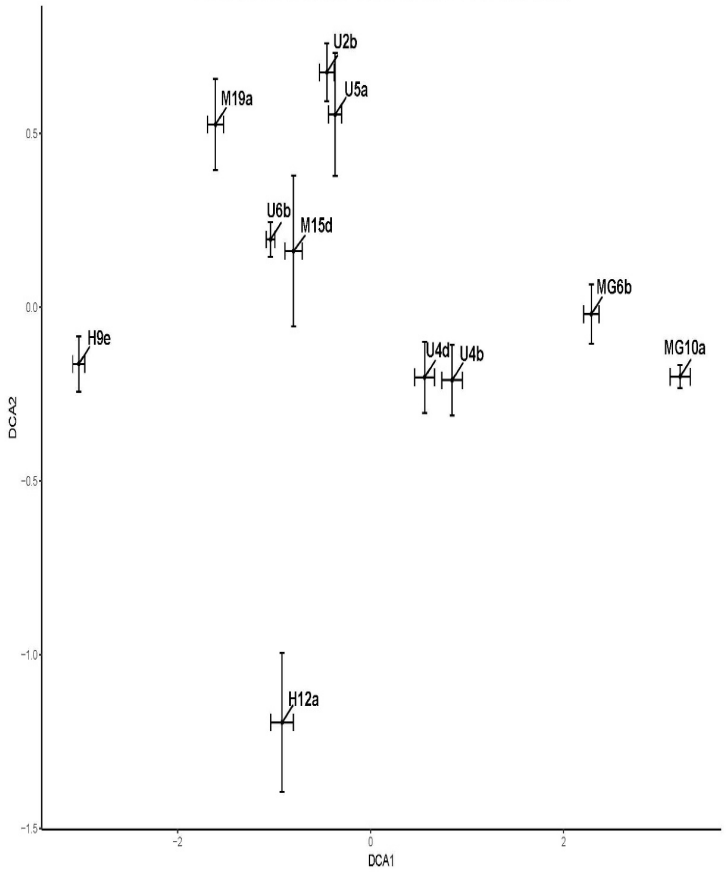


Fig. 3. 4 DCA sub-community mean centroids ± standard error for pseudoquadrats generated from literature (LIT) data for each sub-community type.

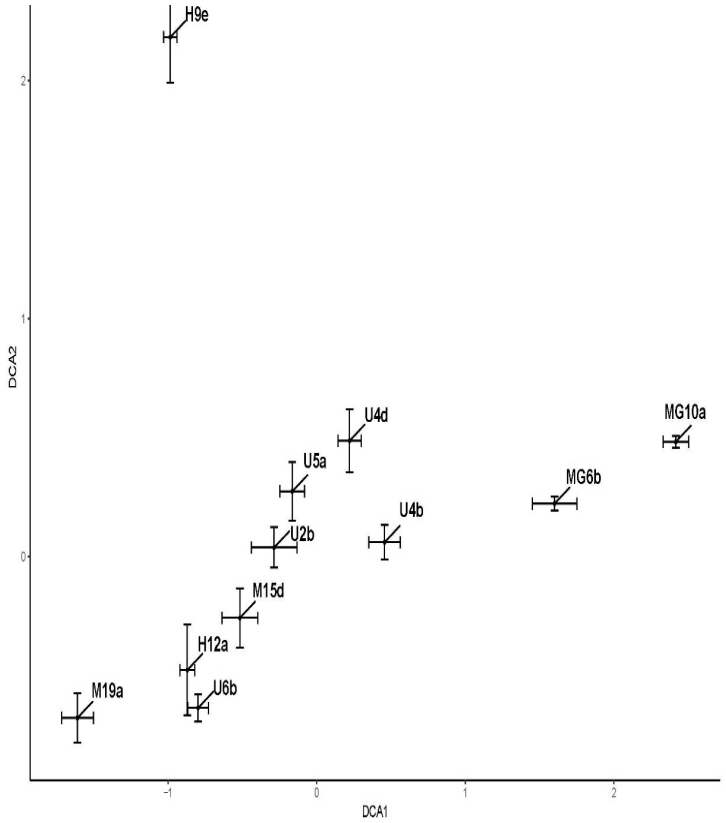


Fig. 3. 3 DCA sub-community mean centroids ± standard error for pseudoquadrats generated from Ashtrees survey data (ASH) for each sub-community type.

3.3.3 Allocate and predict field quadrat data to ordinations (Objective 3)

It is clear from both the LIT and ASH methods (Figs. 3.5 and 3.6 respectively) that many field quadrats were intermediate in species composition, positioned between typical NVC sub-community centroids as defined by the pseudoquadrats. The highest probabilities for each field quadrat were compared to the 'observed' classification (initial TWINSPAN + MAVIS classification). The confusion matrices created to show the numbers of correctly predicted quadrats (Tables 3.2 and 3.3) produced an overall accuracy of 20.36% for the LIT method, and 29.34% for the ASH method, and Kappa values of 0.109 and 0.178 respectively.

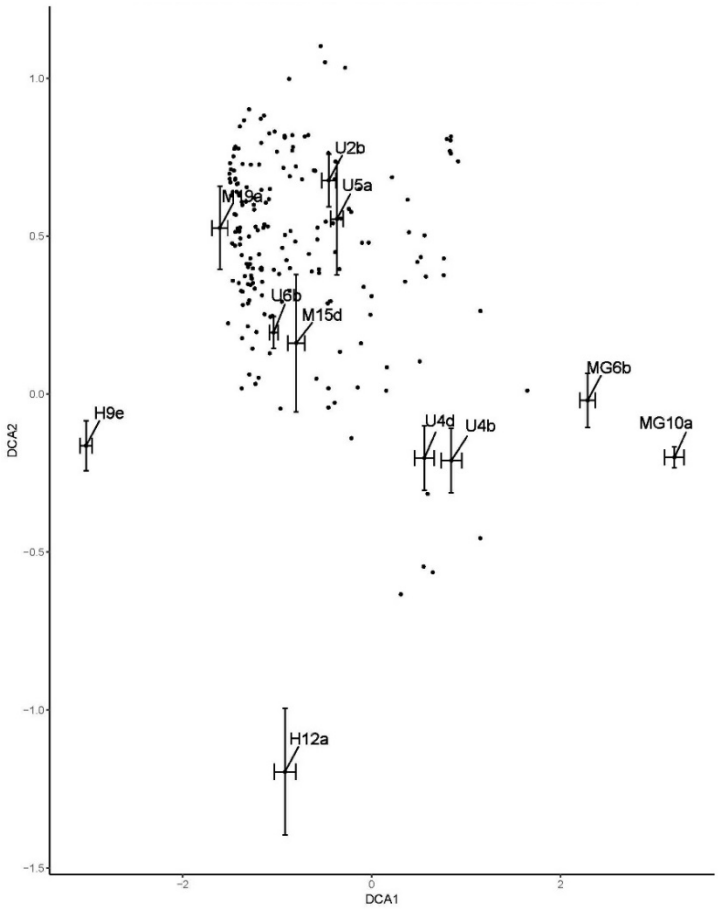


Fig. 3. 5 Sub-community mean centroids ± standard error of pseudoquadrats derived from the NVC handbooks (LIT) in a detrended correspondence analysis. Field quadrats from Ashtrees Dipper (points) displayed as passive samples.

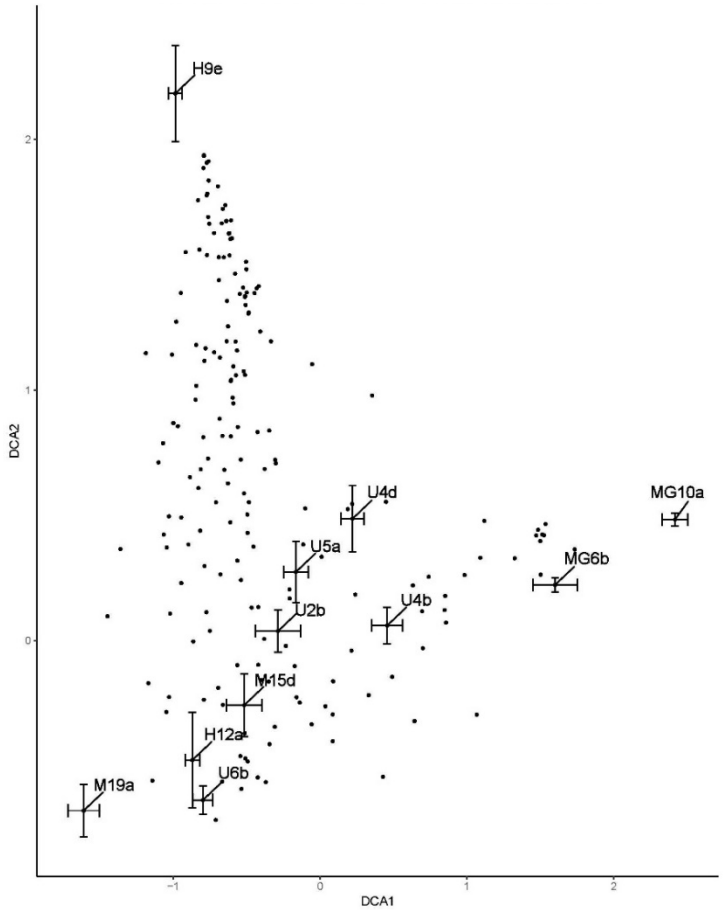


Table 3. 2 Confusion matrix of number of quadrats of highest probability corresponding to same vegetation sub-community classification. Columns: observed communities (MAVIS); rows: predicted via LIT pseudoquadrats.

	H12a	**H9e**	**M15d**	**M19a**	**MG10a**	**MG6b**	**U2b**	**U4b**	**U4d**	**U5a**	**U6b**	**% Total**
H12a	0	0	0	0	0	0	0	0	0	0	0	0.00
H9e	0	0	0	0	0	0	0	0	0	0	0	0.00
M15d	0	0	2	0	0	0	1	0	0	5	2	20.00
M19a	8	18	3	11	2	0	4	0	1	5	1	20.75
MG10a	0	0	0	0	0	0	0	0	0	0	0	0.00
MG6b	0	0	0	0	0	0	0	0	0	0	0	0.00
U2b	0	5	1	1	0	0	3	0	0	7	6	13.04
U4b	0	0	0	0	1	3	4	0	0	2	0	0.00
U4d	0	0	1	0	1	2	0	2	3	3	1	23.08
U5a	0	1	1	0	2	0	5	1	0	8	1	42.11
U6b	2	6	4	1	0	0	9	0	0	10	7	17.95
% Total	0.00	0.00	16.67	84.62	0.00	0.00	11.54	0.00	75.00	20.00	38.89	

Table 3. 3 Confusion matrix of number of quadrats of highest probability corresponding to same vegetation sub-community classification. Columns: observed communities (MAVIS); rows: predicted via ASH pseudoquadrats.

	H12a	H9e	M15d	M19a	MG10a	MG6b	U2b	U4b	U4d	U5a	U6b	% Total
H12a	0	0	0	3	0	0	3	0	0	1	0	0.00
H9e	9	21	2	5	1	0	5	0	0	5	1	42.86
M15d	0	1	2	0	1	0	3	0	0	6	3	12.50
M19a	0	0	0	1	0	0	0	0	0	0	0	100.00
MG10a	0	0	0	0	0	0	0	0	0	0	0	0.00
MG6b	0	0	0	0	1	4	3	0	0	3	0	36.36
U2b	0	0	3	1	0	0	4	1	1	6	4	20.00
U4b	0	0	1	0	2	1	1	2	3	3	3	12.50
U4d	0	1	2	0	0	0	1	0	0	3	0	0.00
U5a	1	7	1	3	1	0	6	0	0	13	5	35.14
U6b	0	0	1	0	0	0	0	0	0	0	2	66.67
% Total	0.00	70.00	16.67	7.69	0.00	80.00	15.38	66.67	0.00	32.50	11.11	

3.3.4 Map community distributions across Ashtrees Dipper (Objective 4)

The resulting interpolation maps show the predicted distribution (via the ASH method) of these communities across the Ashtrees field. The two mires/wet heath subcommuities differed in their distributions: M15d (*Scripus cespitosus – Erica tetralix* wet heath) was relatively common and predicted across much of the site (Fig. 3.7) whereas M19a (*Erica tetralix* subcommunity) was restricted to higher elevations at the southern edge of the site. Two mesotrophic grassland (MG) sub-communities occur at Ashtrees: MG6b (*Anthoxanthum odoratum* sub-community) and MG10a (*Holco-Juncetum effusi typicum;* Typical sub-community). At Ashtrees, neither was abundant, although MG10a had overall lower predictions.

Sub-communities U2b (*Deschampsia flexuosa* - *Vaccinium myrtillus* grassland) and U4b (*Holcus lanatus – Trifolium repens* grassland) ranged across the site, with altitude less important, although mainly in areas of low soil pH (see Appendix 1.2). The U4d sub-community (*Luzula multiflora – Rhytidiadelphus loreus* grassland) had lower predicted abundance, mainly on the eastern side of the site. Similarly, the U6b sub-community (*Carex nigra – Calypogeia trichomanis* sub-community) was sparsely distributed and predicted to be localised to the higher altitudes at the southern edge of the site.

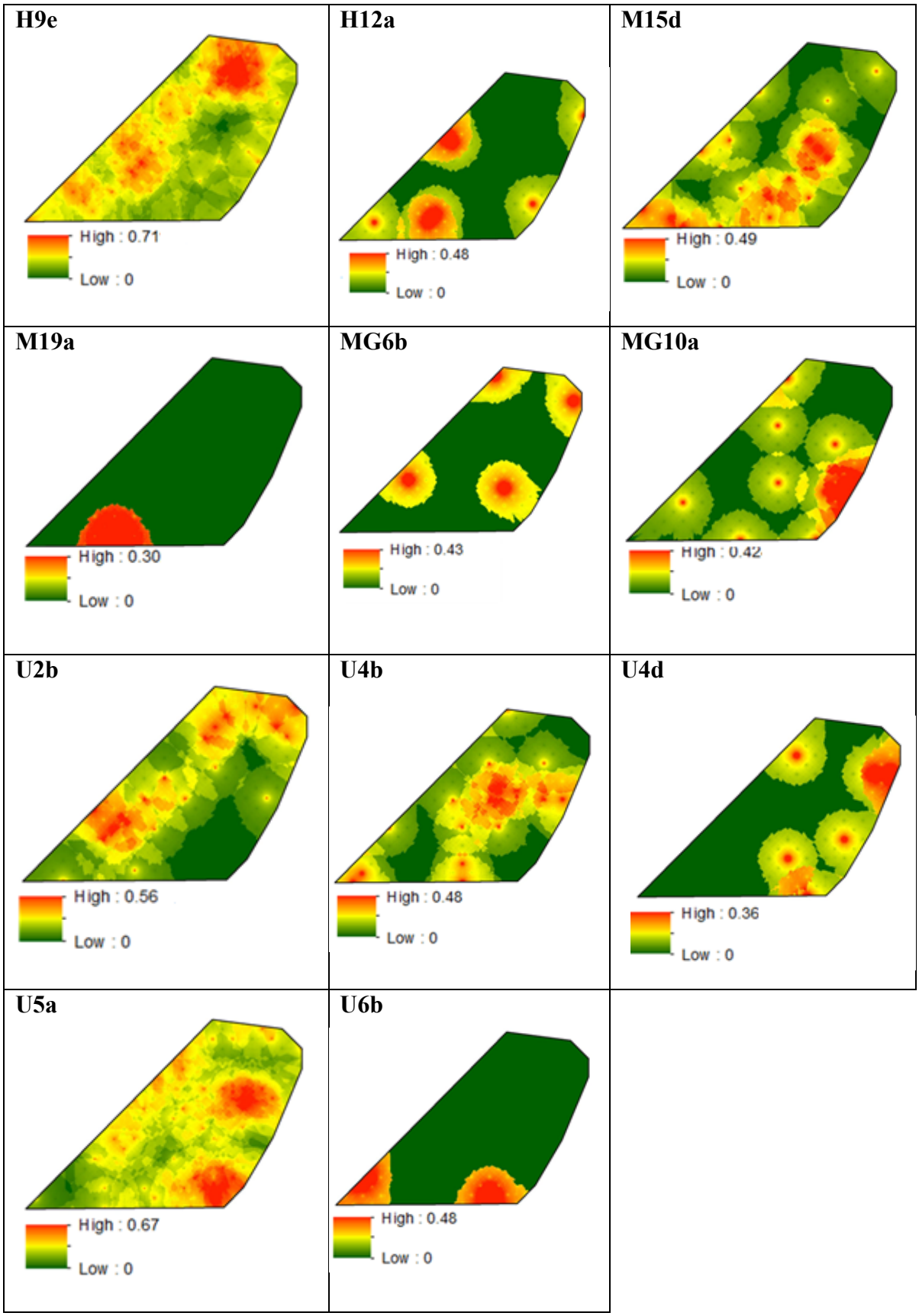


Fig. 3. 6 Interpolated maps of probability of occurrence for the 11 sub-communities occurring at Ashtrees.

3.3 Discussion

This research has demonstrated that both published records and field datasets can be used to generate realised pseudoquadrats that can subsequently be used to allocate quadrats to existing vegetation classification frameworks. Both methods (LIT and ASH) have their strengths and weaknesses, and therefore choice of approach may depend on individual requirements or site characteristics.

3.4.1 Produce an initial "baseline" for NVC classification (Objective 1)

One possible weakness of the TWINSPAN classification used to develop the baseline (Fig. 3.2) is that the number of quadrats per cluster was very variable. Clusters 1, 2, 3 and 5 all contain more than 15 quadrats (clusters 1 and 5 more than 30 quadrats) whereas cluster 7 only contained 3 quadrats. Subsequently after further analysis with MAVIS clusters 1, 6 and 7 were further subdivided. Ideally all the clusters would have contained similar numbers of quadrats to streamline classification between the different community types, but despite fine-tuning several of the TWINSPAN options similar uneven distributions of quadrats between end clusters occurred in all cases. However, it must be remembered that the underlying field survey was based on transects, and not blocks of 'visually homogenous' vegetation, and therefore this type of result is not unsurprising.

3.4.2 Generate "pseudoquadrats" from literature and surveyed data (objective 2)

Fortunately, both the LIT and ASH detrended correspondence analysis ordinations appeared to provide a robust visual summary of the *pseudoquadrat x species* ordination space in two dimensions (Figs 3.3 and 3.4). There was no evidence of an 'arch effect' because DCA is designed to remove this artefact (Hill, 1979). In contrast, in both datasets correspondence analysis (CA) had problems in that they contained a 'wedge' artefact (not shown), suggesting that DCA was the more appropriate analytical method. There appeared to be greater variability in the (sub)community composition when using the LIT method compared (Fig. 3.5) to ASH method (Fig. 3.6) as some of the standard error bars were larger, especially for H12a, M15a and U5a. This might be a result of the LIT method using a larger pool of potential species (i.e. all the species published in the NVC handbook for that sub-community) compared to the more restricted set for the ASH method. In addition, both 'H' and 'U' sub/communities are comprised of very similar species with *Calluna vulgaris*, *Deschampsia*

flexuosa, *Dicranum scoparium* and *Potentilla erecta* amongst the most abundant community constants (Rodwell, 1998a, b, c; JNCC, 2004).

3.4.3 Allocate and predict field quadrat data to ordinations (Objective 3)

The ability to visualise the mean pseudoquadrat communities' positions in ordination space in combination with field samples (as 'passive samples') provides a valuable advantage to the pseudoquadrat approach. It allows users to gain greater insights into how their field quadrats fit into the framework of the classification system they are working with (in this case the NVC), and provides a greater depth of understanding than a mere percentage probability score for an individual field quadrat or group of quadrats.

The greater accuracy of the ASH approach compared to the LIT method was unexpected. While the latter was based on the full set of species described in the standard NVC sub-community descriptions this method would have included numerous species that might not have been at the study site. The greater accuracy of the ASH method, restricting species to those observed at the study site, may simply have arisen because the resultant pseudoquadrats were more representative of species likely to be found in Northumberland, rather than the UK as a whole. This meant that the resultant sub-communities were also more representative local 'variants' of the sub-communities described in the NVC handbooks. It should also be noted that the low accuracy of this method, compared to MAVIS, might have stemmed from the fact that numerous communities could, in reality, overlap in their classification. It should be noted that the probability scores obtained for each community (from both LIT and ASH methods) always resulted in one highest score (i.e. a passive sample always had one highest community score) while the MAVIS method, in some cases, had multiple identical matching coefficients.

At Ashtrees it appeared that the vegetation communities were more species-poor when compared to the species listed for these communities in the British NVC handbooks. Whilst this might partly have been attributable to the use of smaller quadrats at Ashtrees (1m rather than the NVC standard 2m, for such habitats), it is unlikely to be a major factor, especially given the large number of quadrats used in the survey. The 10km dot-distribution maps of several sub-communities in the NVC indicate that parts of North East England were not available for those sub-communities in the original survey (pre-1991) when the NVC was constructed, compared to the updated dot-distribution maps from 2015. As such, the

published sub-community descriptions may not necessarily accurately describe the species likely to be encountered in those same sub-communities in certain parts of the UK.

One possible improvement to predicting community probabilities based on the position of passive samples in the pseudoquadrat ordination space would be to up-weight communities with large standard error bars. For example, H12a, H15a and U5a had relatively large SE's for the LIT method, therefore the predicted probabilities for a field passive sample falling near these sub-communities could be up-weighted. Conversely MG10a and U6b had relatively small SE bars (probably reflecting more narrowly-defined communities in terms of their species) and therefore could be down-weighted. Both the ASH and LIT methods of generating pseudoquadrats have advantages and disadvantages. The ASH method is effectively 'site-specific' in that the list of species used reflects that observed at any individual study site. However, this may mean that it is more difficult to interpret the results in the context of an existing national classification system. The LIT method uses a national system (here the NVC) and so is more generalisable, but is therefore at greater risk of including species which might not occur locally at a study site or might not be detected if at very low constancy. Choice of approach depends on the requirements of the user. The key advantage of both methods is that they can be used on individual quadrats, including those obtained by other sampling methods such as transects (as here), and not restricted to 'homogenous vegetation' or a minimum number of quadrats.

3.4.4 Map community distributions across Ashtrees Dipper (Objective 4)

Some vegetation communities, e.g. H9e, M15d, U2b, U4b and U5a had relatively high predicted distribution patterns across much of the Asthrees heft. This could mean that at smaller spatial scales, the range of vegetation communities might be neighbouring or overlapping. This largely depends on the variation in vegetation composition of each communities, given that some of the communities identified at Ashtrees (e.g. U2b, U6b, U4a etc.) have similar species composition, or are relatively species poor. The habitat requirement for U2b (*Vaccinium myrtillus* sub-community), U4a (Typical) and U6b (*Carex nigra-Calypogeia trichomanis* sub-community) are similar in that they typically grow in moist peaty soils as well as preferring cooler humid climates, as is found in the NE England. Grazing is important in maintaining these communities especially in locations of rough-grazing schemes that allow the community to grow on slopes and mid- to high-altitudes.

Calluna vulgaris is abundant in H9e and therefore the community's wide predicted distribution was to be expected since Ashtrees is abundant in heather especially at higher altitudes where drainage ditches were introduced. It was predicted to be less frequent in eastern parts of the heft, where soil water content is lower, soil pH is more acidic (Appendix 1.1) and sheep grazing is higher. These eastern areas were predicted to be better for mesotrophic grassland MG10a, and acid grassland U4d, possibly reflecting soil and grazing characteristic. Some communities were predicted to occur only relatively rarely across the heft, especially M19a and U6b (Fig. 3.7) and it is more difficult to identify possible environmental or management factors that might drive their distribution. These two communities can have substantially different vegetation species within their assemblages and therefore their 'isolated' distribution can be inferred from the similar habitat requirements. M19a, a blanket bog and U6b, a shade-sensitive hydrophilic sub-community require both require high altitudes where and wet environments. Gentle slopes at higher altitudes at Ashtrees can explain the distribution of these communities in these areas.

This interpolation approach is valuable as it demonstrates that it is possible to generate high spatial resolution map (here 1m resolution raster) from much coarser survey data (75m x 150m quadrats along transects). While validation, at such small spatial scales, requires additional field surveys, the high number of quadrats in close proximity to each other suggest reliable community distribution results. Such maps have considerable potential value for site managers, especially to identify possibly unsurveyed areas in which to target conservation management.

3.5 Conclusion

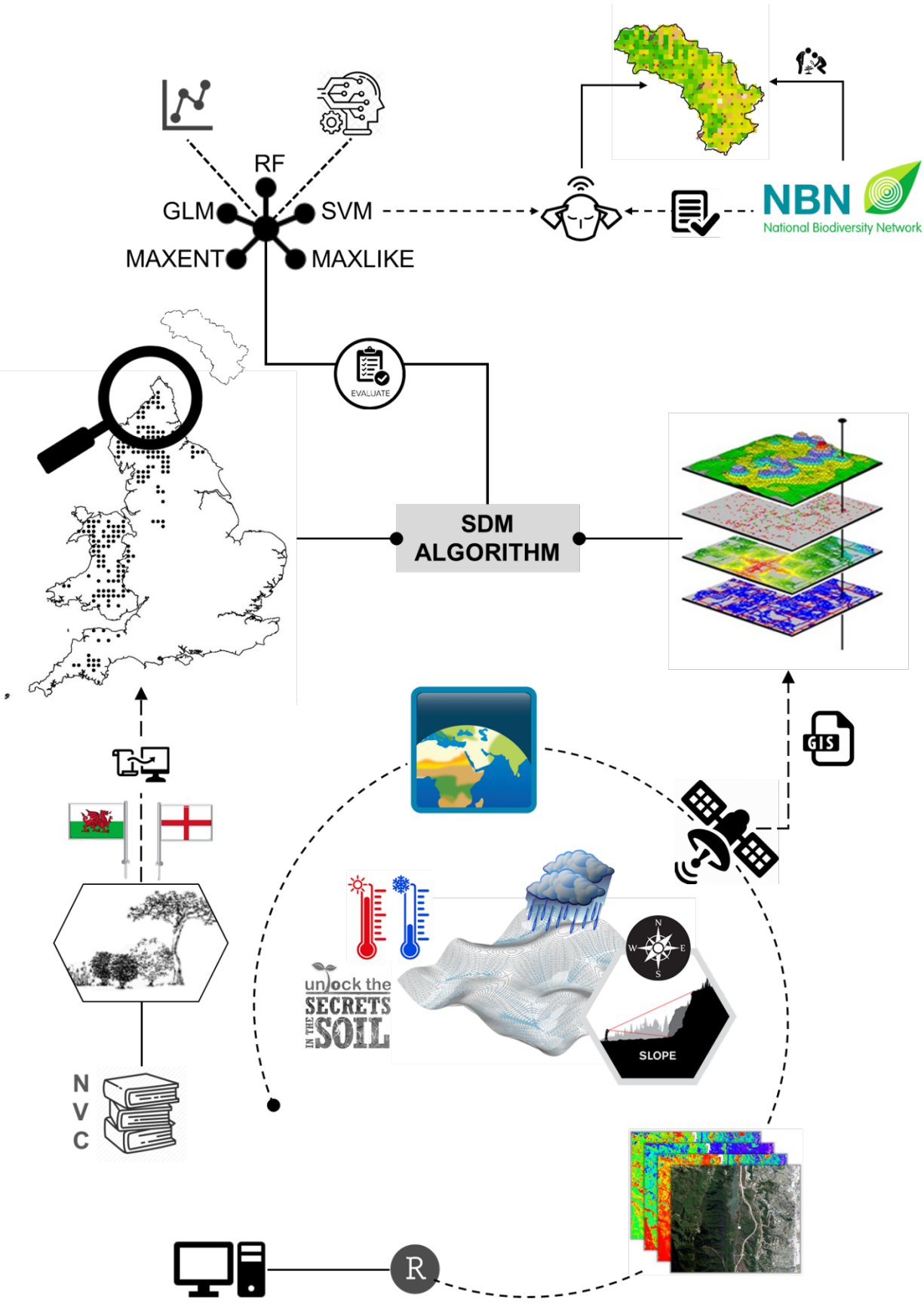
Pseudoquadrats can provide a reliable method to existing techniques to allocate quadrats into extant vegetation classifications. The techniques described in this chapter are simple, and could be easily modified to account for the characteristics of any national or regional classification. The use of TWINSPAN as an initial classification technique is an important tool for vegetation classification. In cases where data was not gathered through the conventional methods, hierarchical classification clusters quadrats of similar species composition together. This can then be used as a baseline technique to generate vegetation communities, especially in scenarios where classification software is unavailable. The pseudoquadrat techniques can create vegetation communities in ordination space, in which vegetation sample data can then be inputted into the ordination as 'passive' samples. The measure of distance between the 'passive' samples and the pre-generated community centroids provides an indication of the community (or sub-community) to which the sample data belongs. The probability data obtained for each community within a sample can then be interpolated across any given field, identifying areas where these vegetation communities are most likely to occur.

A potential improvement in this classification method would be to classify vegetation using the same technique but on additional, similar upland sites. This would possibly require adding more pseudoquadrat centroids of different vegetation communities in the ordination methods. An ordination for each broad class type (e.g. 'H', 'M', 'MG', 'U' etc.) can be developed. This will allow a readily-available system to input percentage cover data from any vegetation sampling as passive quadrats for a quick classification. One weakness that this system has is that very different vegetation communities, for example, woodland, grassland and sand dune habitats, would not provide a good ordination plot because of the high variance between the different sample types (and centroids).

Some national vegetation classifications provide abiotic data associated for each vegetation class. This could be used in combination with, for example GIS maps etc., or published maps of species-distributions, to fine-tune the construction of pseudoquadrats depending on the end-user's requirements. The flexibility in customising the construction of pseudoquadrats, visualising them in ordination space, and using them to make predictions for new field quadrats provides the developed methods with their attraction.

Chapter 4. Prediction of vegetation community distribution across England and Wales

Graphical abstract



Abstract

The current increase in availability of open source remote-sensed data makes prediction of species distributions and community composition across a wide range of spatial scales more practical, especially for vegetation. Drone technologies can survey small areas of vegetation down to centimetre-level spatial scale, whilst satellite data can measure vegetation across larger areas, albeit at coarser spatial resolution. Furthermore, multispectral imagery improves the characterisation of the vegetation, for example via NDVI and EVI. Multi-spectral remote-sensed data is becoming an important addition for both ecological assessment of extant vegetation, and prediction of species, due to increases in both its spatial and temporal resolution, especially when used in conjunction with other sources of environmental and meteorological data. Species distribution models (SDMs) are empirical models that have been developed to predict species occurrence from presence-data and environmental variables and can utilise a range of different underlying model techniques. SDMs are not restricted to predicting species but can also predict communities if the presence-data on the spatial distribution of the latter is known. Plant communities have different requirements in terms of the abiotic and biotic environment in which they occur, and these must be incorporated into SDMs in order to make realistic predictions.

High-resolution digital maps of environmental variables including topography, meteorology, soil type, land cover and enhanced vegetation index (EVI) were used to predict NVC communities that occur within the upland priority habitats in the UK Biodiversity Action Plan (UK BAP). Published NVC 10km resolution distribution maps, derived from those in the NVC handbooks, were used as the 'observed' presence-data of community distributions to create SDMs to predict community distribution maps for the whole of England and Wales at 1km resolution. SDMs based on five different modelling approaches were investigated: generalised linear models (GLMs), support vector machines (SVM), random forests (RF), maximum entropy (MaxEnt) and maximum likelihood (MaxLike). Model accuracy was assessed using measures for area under the curve (AUC) and true skill statistic (TSS), and the RF approach was consistently the most reliable. After prediction of NVC communities, constancy level III, IV and V species from the NVC accounts were identified and their distribution extracted from the NBN Atlas. These observed species distributions were compared to the NVC predictions, using the catchment of the River Rede in Northumberland, UK, as a case study area.

This chapter shows that vegetation communities can be reliably predicted at very large spatial scales using species distribution models. The resulting community distribution maps of threatened and/or prioritised vegetation communities provide a sound basis to predict both the location of communities and their constituent species for sampling and analysis by both professionals and citizen scientists. The results can help scientists and volunteers in identifying possible locations of communities and species of interest which can help add data to both the NBN Atlas and the National Plant Monitoring Scheme (NPMS).

4.1 Introduction

Vegetation community assessment has traditionally depended on initial collection of data through field surveys of vegetation, and classification of communities can then be performed through a number of techniques as described in Chapter 3. Vegetation mapping and prediction is a very important tool to provide information and identify sites that are of high conservation value (Hearn *et al.*, 2011). Vegetation community surveys were carried out across Great Britain in the development of the NVC, and results are detailed in the 'British Plant Communities' handbook series (Rodwell, 1998a; Rodwell, 1998b; Rodwell, 1998c). The handbooks include maps of the 10 km squares in which samples for each community were recorded as part of the development of the NVC, and these can be considered as 'presence-only' maps (Franklin, 2010), i.e. the absence of a record for a community in part of the map does not necessarily indicate that the community might not occur at that location. There may also have been climatic, management or other environmental changes since the NVC maps were published, that might affect the distribution of the NVC communities nationally. The NVC 10km distribution maps are now available in digital format for most NVC community types via Joint Nature Conservation Committee (JNCC, 2016).

4.1.1 Species Distribution Models

Species distribution models (SDMs) are empirical models to predict species occurrence, using environmental variables. After creation of an SDM, the resulting data can then be used to predict distributions (Liu *et al.*, 2009; Liu *et al.*, 2011). SDMs have been used to understand the spatio-temporal ecological processes that affect populations or communities with changing climate and/or species invasion (Franklin, 2010; Fournier *et al.*, 2017). The recent improvement in SDMs opens opportunities for more flexible and reliable modelling in ecology (Rushton *et al.*, 2004). Environmental heterogeneity and change occurs across multiple scales and therefore prediction of species distributions can be challenging (Rushton *et al.*, 2004; Franklin, 2010). Species presence data was a primary requirement for understanding and predicting species distributions and predictions across multiple scales (Guisan *et al.*, 2006; Warton *et al.*, 2016). Modelling techniques, e.g. generalised linear models (GLMs), generalised linear mixed models (GLMM) or general additive models (GAMs) have been used to predict the distribution of species (Warton *et al.*, 2015). Machine learning techniques such as random forest models (RF), generalised dissimilarity models

(GDMs) and support vector machine (SVM) have recently gained popularity to fit an array of distribution models with reported increased reliability on their predictions (Franklin, 2010; Naimi and Araújo, 2016). It is, therefore, important to select realistic predictor variables, to account for the likely physiological requirements of the target species (Rushton *et al.*, 2004; Austin, 2007; Austin and Van Niel, 2011). The same predictor importance applies for the use of 'ensemble' forecasting models whereby the models can be 'weighted' depending on the preferred technique (e.g. GLM, SVM, RF etc.). Ensemble forecasting uses numerous modelling techniques to produce one model output based on the accuracy of each individual model produced. This is done using 'weighted' averages between models, giving a higher weighting to more accurate models and a lower weighting to less accurate models (Franklin, 2010; Naimi and Araújo, 2016; Fournier et al., 2017). This research had originally considered using 'ensemble' models, but as a key aim was to compare different modelling techniques, it was decided to use individual model techniques for individual comparison of prediction and accuracy.

SDMs depend on multiple environmental (abiotic and biotic) variables to predict species distributions (Franklin, 2010; Fournier *et al.*, 2017) and restricting the number of predictors to the most effective ones maximises the performance of SDMs and the accuracy of the models (Fournier *et al.*, 2017). The predictor variables act as filters to determine the species distribution (Blach-Overgaard *et al.*, 2010; Araújo and Peterson, 2012; Fournier *et al.*, 2017). With increased availability of remote-sensed data plus integration with geographic information systems (GIS) management of spatial data is much simpler (Rushton *et al.*, 2004), and makes SDMs easier to implement. High quality digital maps of environmental variables permit interpolation and extrapolation across larger areas (Franklin, 2010), making it feasible to predict species distributions in areas where, for example, it is difficult to collect field data (Rushton *et al.*, 2004; Smith and Wyatt, 2007). Note that it is important to first validate the models by additional techniques, for example field surveys, model simulations or machine learning techniques, before fully relying on the model predictions.

4.1.2 Evaluation and assessment

While prediction of species distributions is important for both pure and applied ecologists, it is imperative that the resulting models are evaluated for accuracy (Guisan and Thuiller, 2005; Guisan *et al.*, 2006; Liu *et al.*, 2009; Liu *et al.*, 2011; Naimi and Araújo, 2016). The type of

evaluation best suited for the models depends on the training data used, e.g. presence-absence, presence-only and presence-background. The latter is gaining popularity since species data is, in most cases, presence-only data, and can streamline and increase reliability of the model outcomes (Jiménez-Valverde, 2012). Background data are randomly selected from the area of study, finding functions within the datasets to discriminate between locations of species presence and highly-probable absences (Jiménez-Valverde, 2012). Evaluation of accuracy can be of two types: threshold-dependent or threshold-independent (Manel *et al.*, 2001; Guisan and Thuiller, 2005; Elith *et al.*, 2006; Allouche *et al.*, 2006; Liu *et al.*, 2009). Two main types of model accuracy evaluators are used to assess accuracy of SDMs: AUC: the area under the receiver operating characteristic (ROC; Allouche *et al.*, 2006; Liu *et al.*, 2009; Franklin, 2010; Naimi and Araújo, 2016), and the true skill statistic (TSS; Allouche *et al.*, 2006; Liu *et al.*, 2011; Eskildsen *et al.*, 2013). AUC is threshold-independent and takes into account the proportion of locations that are occupied by the species (Hanberry and He, 2013; Lawson *et al.*, 2013). Cohen's Kappa and TSS are threshold-dependent, in that they depend on a threshold probability above which predicted species are present. In generalised SDMs, sensitivity and specificity (Franklin, 2010; Jiménez-Valverde, 2012) have also been traditionally used but are not used as individual evaluators in this research. Both AUC and TSS are based on the ratio (either positive likelihood ratio or negative likelihood ratio) between sensitivity and specificity (Allouche *et al.*, 2006). Sensitivity is the probability that the SDM test result is positive when the species or community (in this case) is present (Allouche *et al.*, 2006). Specificity is the probability that the SDM test result is negative when the species or community is absent (Allouche *et al.*, 2006). Unlike Cohen's Kappa, AUC and TSS operate independently of bias (also referred to as 'prevalence') between observations and the distribution of data across the SDM predictors (De Marco and Nóbrega, 2018).

ROC plots have been adopted as standard methods for assessing accuracy of SDMs.
An AUC higher than 0.5 implies higher accuracy of the model. However, this is a general measure in models that have been trained with presence-absence data. Using presence-background data can shift the 0.5 threshold, making the AUC measure less reliable for this type of training data (Allouche *et al.*, 2006; Freeman and Moisen, 2008; Jiménez-Valverde, 2012). To counteract this uncertainty, Cohen's Kappa and TSS, both ranging from -1 to +1, are good supplementary assessors of models. TSS takes into consideration the dependence of Kappa on prevalence while still pertaining the benefits of the measure of performance and avoids the need of a single threshold (Guisan and Zimmermann, 2000; Manel *et al.*, 2001; Allouche *et al.*, 2006; Freeman and Moisen, 2008; Liu *et al.*, 2009; Franklin, 2010). TSS also

accounts for both sensitivity and specificity (1-max{sensitivity + specificity}) of the model and is thus a better suited standard for evaluation of SDM models (Allouche *et al.*, 2006; Franklin, 2010).

4.1.3 Environmental drivers and predictors

Species-environment relationships are scale-dependent, for example, soil pH or water content might be important at a 1m spatial scale, whereas altitude or soil type would be more important at a 1km scale. Understanding these relationships requires continual development in biological modelling as tools to aid conservation and decision-makers. This research looks into the possibility of using SDMs to predict vegetation communities rather than species, to provide an objective method on where vegetation communities are more likely to occur. Franklin (2010) outlines the concept of landscape modelling of plant community dynamics for predictive modelling on potential habitats of occurrence. While Franklin (2010) discusses this in terms of climate and anthropogenic influence, SDMs can be modified to encompass multi-scale environmental simulations.

This chapter makes use of high-resolution, large-scale spatial and spatio-temporal environmental data obtained from a range of different sources. As Franklin (2010) and Bradter *et al.*, (2011) note, climatic data is an essential resource for appropriate use of vegetation distribution modelling. Different plant communities have different requirements of and tolerances to different abiotic and biotic systems. To encompass the major ecological requirements, this research uses high resolution digital maps of elevation (digital elevation models; DEMs), slope, aspect, soil type, seasonal maximum and minimum temperature, annual rainfall (precipitation), landcover and satellite data to predict selected vegetation communities across England and Wales. This research limits vegetation community distribution prediction to those currently deemed to require conservation efforts under the UK Biodiversity Action Plan (UK BAP).

4.1.4 UK BAP and other policies

The UK Biodiversity Action Plan (UK BAP; JNCC, 2016) describes the biological resources of the United Kingdom and provides action plans for the most threatened species and habitats that required higher levels of management and recovery. The UK BAP has been updated in

2012 to the UK Post-2010 Biodiversity Framework (JNCC, 2019), indicating changes and targets in conservation networks to be in place by 2020. This action plan is part of the requirement under a range of European Directives and legislations to conserve biodiversity through the Ecosystem Approach outlined by the Conservation of Biological Diversity (CBD; United Nations, 1992). Within these frameworks ecosystem structure and functioning are given due importance both across time (long-term planning) and space (e.g. protected areas) at all ecological levels. While the UK BAP details conservation of species and habitats founds within broad land types, vegetation communities are amongst the most important because their role in defining the habitat (DEFRA, 2007). Since gaps are present in published vegetation community distribution maps due to the high sampling effort required, especially on a national scale, reliable prediction of vegetation communities, and their constituent species, will reduce the need for additional new sampling, while providing data for policy makers to indicate potential sites for additional conservation and management.

The European Habitats Directive (Council Directive 92/43/EEC) and the Birds Directive (Directive 2009/147/EC) serve as guided legislation on habitat conservation, prioritising habitats as part of the NATURA 2000 scheme that are of importance for multiple animal and vegetation species across threatened landscapes (European Commission, 1992; European Commission, 2009; McLeod *et al.*, 2009). The Habitats Directive operates to sustain and protect environmental landscapes at any given spatial scale and their inhabitants. The rarity of certain vegetation communities and their constituents is amongst such prioritisation and additionally these habitats link different policies. The Birds Directive is one such link, especially in deteriorating upland environments, ensuring that protection of vegetation communities will allow birds and other such animals to thrive in land cover types that are influenced by vegetation communities and resulting landscape as the primary building blocks.

4.1.5 The National Biodiversity Network Atlas (NBN Atlas)

The National Biodiversity Network Atlas is an open source online user interface tool and database that provide a platform for innovation of citizen science. The main aim of this is to improve biodiversity knowledge about the natural world and help the scientific community in managing gaps in species presence data (Ryan, 2018; National Biodiversity Network, 2019). The NBN Atlas was developed to provide an effective national data management system for reporting of species occurrences (Lawrence, 2010). This can reduce survey efforts and

produce a suitable database that provides information on location, presence, distribution and spatio-temporal data of different species (NBN Atlas). Planning regulations for conservation and management can then be adopted by scientists and policy makers to ensure proper policy evaluations (Pollet *et al.*, 2001; Lawrence, 2006; Lawrence, 2010). The NBN Atlas also provides scope for biodiversity monitoring through citizen science. This concept can be applied to different countries, where details of species distribution can be shared though an online database and can be of use in proper management (Pollet *et al.*, 2001). While gaps in species presence and distribution are difficult to overcome, the use of this tool in wildlife management and prediction can be extended to population and community ecology at large-scale studies (Callcutt *et al.*, 2018). The continuous recording of species occurrence data provides an invaluable tool for spatio-temporal species and community prediction of distribution and status change (Callcut *et al.*, 2018; Ryan, 2018). The NBN Atlas, in this research, is used to 'verify' vegetation community distribution predictions.

4.1.6 Aims and Objectives

The overall aim of this chapter is to develop a robust system to predict the probability of occurrence of NVC communities across England and Wales, and compare with the observed incidence of their most common constituent species, occurring at NVC constancies III, IV and V.

The specific objectives are to:

1. Collate all relevant environmental data that may affect plant community distributions, including soil type, altitude, slope, aspect, meteorology, land cover and raw satellite data;
2. Randomly select a set of 10 target NVC communities from the UK BAP that occur in the uplands of England and Wales, to be used with the 10 communities already identified as occurring at Ashtrees Dipper in the Rede Catchment (Chapter 3);
3. Develop SDMs for the 10 communities randomly selected from UK BAP, plus SDMs for the 10 Ashtrees communities, for the whole of England and Wales, and evaluate them using AUC and TSS
4. Identify the most frequent species (based on their published constancy scores) in these two sets of 10 communities. Compare the observed distributions of these species in the NBN Atlas with predicted NVC community scores, using the catchment of the River Rede in Northumberland as a case study region.

4.2 Methodology

4.2.1 Collation of predictor data for use in SDMs (Objective 1)

England and Wales was used to predict NVC communities at a large scale (Fig. 4.1). The borders of England and Wales were obtained as shapefiles (.shp) under OS open data terms and conditions provided by Ordanance Survey (2019) and provided a 'clipping' boundary for other data used. All subsequent feature and raster data were all converted from the original coordinate reference system (CRS) to the British National Grid coordinate reference system (EPSG:27700) and feature data manipulated with the 'sf' package (Pebesma *et al.*, 2019) and raster data the 'raster' package (Hijmans *et al.*, 2019) in R (R Core Team, 2013). All maps were converted to 1km resolution raster maps by nearest neighbour resampling prior to use in the SDMs.

Fig. 4. 1 Border of England and Wales. Data obtained from Ordanance Survey, (2019) https://www.ordnancesurvey.co.uk/opendatadownload/products.html.

2015 Land cover (LCM2015)

The Centre for Ecology and Hydrology (CEH) Land Cover map 2015 (LCM2015; Rowland *et al.*, 2017) at 25m resolution was obtained as a TIFF raster from EDINA Digimap Environment Data Download (https://digimap.edina.ac.uk/roam/download/environment). The LCM2015 is a parcel-based land cover map for the United Kingdom that classifies land cover in 10 aggregate classes, 10 broad habitat types and 21 target classes. For the purpose of this research, the LCM2015 was clipped to England and Wales and only the target classes are used in analysis.

Topography

Digital Elevation Model (DEM) GEOTIFF raster digital maps were obtained from U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center as ASTER Global DEM (USGS, 2010). The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) is a NASA radiometer that collects global stereo near-infrared digital elevation data from a NASA spacecraft (ASTER GDEM Validation Team, 2011). This makes data highly accurate with DEM data collected at 1 arc-second resolution (approx. 30m resolution) and are freely available in tile format and WGS84 ellipsoid from Earth Explorer online user interface (USGS, 2010; https://earthexplorer.usgs.gov/). Slope and Aspect across England and Wales were computed at 30 m resolution from digital elevation data via the 'terrain' function (below) from the 'raster' package (Hijmans *et al.*, 2019). Both were computed in degrees: slope (0^{o} to $< 90^{o}$) and aspect (0^{o} to 360^{o}).

Meteorology

Monthly minimum and maximum temperature data for years ranging 1988 to 1992 was obtained from the CEDA Archive available from the MET Office (STFC, 2018, https://catalogue.ceda.ac.uk/uuid/4dc8450d889a491ebb20e724debe2dfb). The dataset is available as a 1km resolution, HadUK-Grid grdded and regional average climate observation for the UK. This range of temperatures coincides with the temporal period when the National Vegetation Survey was conducted as detailed in the NVC handbooks (Rodwell, 1998a). Mean minimum and mean maximum annual winter (January, February and December) and summer (June, July and August) temperature data between 1988 and 1992 was calculated, to produce four raster maps, each covering a 5-year range. Annual rainfall data for years between 1988 and 1992 was obtained from MET Office (2018) at 1km resolution. Mean annual rainfall between 1988 and 1992 was calculated and projected for England and Wales.

Soil data (NATMAP)

National soils data for England and Wales was obtained under license from the National Soil Resources Institute (NSRI) at Cranfield University, United Kingdom (Cranfield University, 2017). The NATMAPsoilscapes (1km resolution) digital map represents the major soil types across England and Wales and shows geographic soils associations identified by frequently occurring soils with predominant major soil types grouped together (Cranfield University, 2017). The soil types are differentiated by observable or measurable characteristics of the sampled soil profile.

LANDSAT satellite data

Remote-sensed LANDSAT 8 OLI (Operational Land Imager)/TIRS(Thermal Infrared Sensor) satellite data was obtained from USGS using the Earth Explorer online user interface for England and Wales (USGS, 2019; https://earthexplorer.usgs.gov/). LANDSAT 8 OLI/TIRS provides surface reflectance data at 30m resolution that includes, in downloadable format, 7 bands of surface reflectance data as well as level 2 Pixel Quality Band (pixel_qa), Aerosol QA (sr_aerosol) and Radiometric Saturation QA (radsat_qa) that provide global landmass coverage in visible, Near Infrared (NIR) and SWIR (Short Wave Infrared) Imagery (U.S. Geological Survey, Earth Resources Observation And Science Center, 2012; Masek, 2019). The Operational Land Manager (OLI) collects data for coastal and cirrus band as well as LANDSAT multispectral bands, with bandwidths being refined for 6 'heritage' bands (the original 6 bands) The Thermal Infrared Sensor (TIRS) collects data for narrow spectral bands that fall in the thermal regions previously collected by LANDSATs 4-7 (NASA, 2019).

Enhanced Vegetation Index (EVI)

The different multispectral bands allows derivation of additional spectral index products, e.g. Normalised Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI; U.S. Geological Survey, Earth Resources Observation And Science Center, 2012). LANDSAT 8 mosaic tiles were obtained, in compressed format, for England and Wales, ensuring that each tile has minimal cloud coverage. The LANDSAT 8 mosaic tiles were reprojected to the British National Grid coordinate system (EPSG: 27700) and merged together, ensuring that extent of each mosaic tile was set appropriately, and if mosaic tiles overlapped the best mosaic used (i.e. with little or no cloud coverage). This was repeated for each useable colour band (red, blue, near infra-red). The resulting England-Wales merged bands were stacked using the 'stack' and 'brick' functions from the 'raster' package (Hijmans *et al.*, 2019). The

Enhanced Vegetation Index was derived using the 'LSRS' package (Sarparast, 2018) within the R statistical software (R Core Team, 2014). The resulting EVI images where then clipped to the boundaries of England and Wales and merged together as a mosaic raster.

4.2.2 Select target NVC communities and collate distribution map data (Objective 2)

The NVC vegetation communities highlighted in UK BAP as conservation importance in the following broad headings were selected, but restricted to uplands (over 250 m): calcicolous grasslands (CG), heaths (H), mires (M), calcifugous and montane grasslands (U), woodland and scrub (W) and mesotrophic grasslands (MG). Fourty-four communities were identified as meeting these criteria, and 10 were selected at random for detailed further study. A second set of 10 communities, derived from the analyses in Chapter 3, were use in similar studies (Objectives 3 and 4). The dot-distribution maps for the selected 20 NVC communities involved in this chapter are provided in Appendix 2. Vegetation community data used in this research chapter are available as dot-distribution maps originally published in the 'British Plant Communities' handbook (Rodwell, 1998a, b, c) series, an example shown in Fig. 4.2 for the H12 community. They are now freely available as an interactive Microsoft Excel maps, via the Upland NVC types mapping tool (version 2) on the JNCC website (JNCC, 2009; http://jncc.defra.gov.uk/page-4267) for Great Britain. The vegetation communities assessed in this chapter are those requiring conservation under the UK Biodiversity Action Plan (UK BAP) and EU Habitats Directive, summarised by Jackson and Gaston (2008) in JNCC, (2016), Report 307 from the JNCC report series (http://jncc.defra.gov.uk/page-2433#1401) and Biron (2010). A number of vegetation communities that are part of the UK BAP were selected (Table 4.1). Data were exported from the downloaded Microsoft Excel files and imported into R using the 'sf' package (Pebesma *et al.*, 2019) in R statistical software (R Core Team, 2014).

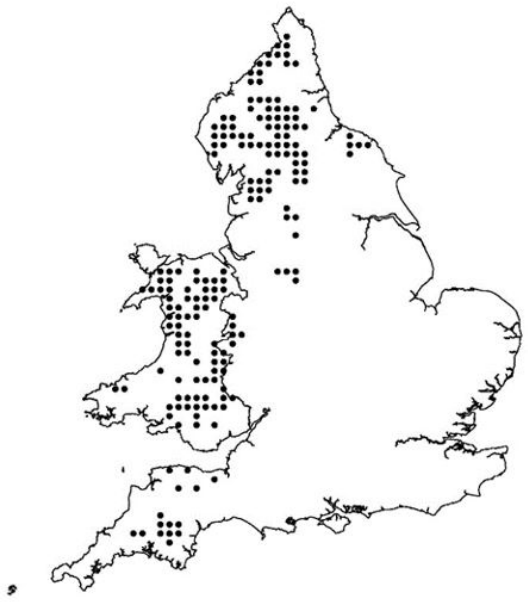


Fig. 4. 2 Dot-distribution map of H12 surveyed in England and Wales. Adapted from the Upland NVC types mapping tool (version 2), originally published in Averis, A., Averis, B., Birks, J., Horsfield, D., Thompson, D. & Yeo, M. 2004. An Illustrated Guide to British Vegetation. Peterborough, Joint Nature Conservation Committee.

4.2.3 Development and testing of NVC species distribution models (Objective 3)

Species distribution models were generated for each selected NVC vegetation community. Species distribution modelling is a multi-step process with each step using either the 'raster' (Hijmans *et al.*, 2019) 'sf' (Pebesma *et* al,, 2019) and 'SDM' (Naimi and Araújo, 2019) packages in R statistical software (R Core Team, 2014) depending on the requirement. Vegetation community distribution presence data that were previously obtained from the JNCC website (JNCC, 2009) and converted to shapefiles were converted to a spatial dataframe object, using the 'sf' package (Pebesma *et al.*, 2019). An R dataframe for all the observed NVC communities was created with four columns: presence of community (all values set to 1), easting, northing, community code name. The environmental predictors rasters were stored as a 'raster stack' which is a set of raster layers with the same spatial extent and resolution were combined as one raster object (Hijmans *et al.*, 2019). The vegetation community data is presence-only data therefore 100 random pseudoabsence datapoints (or background data) were generated for each community. Some communities had relatively few records (see Table 4.1) therefore it would have been unrealistic to have set a

high number of pseudoabsence records. After generation of the pseudoabsence records, these were combined with the observed records and the raster stack of environmental (predictor) data to create an 'sdmData' object, for use in the SDM itself. In addition, the first step in creating the SDM also included a training (80%) and testing (20%) split of the data to increase reliability of the output. The 'trained' model was then used in the second part of SDM creation. A flowchart summarising the procedure is given in Fig. 4.3.

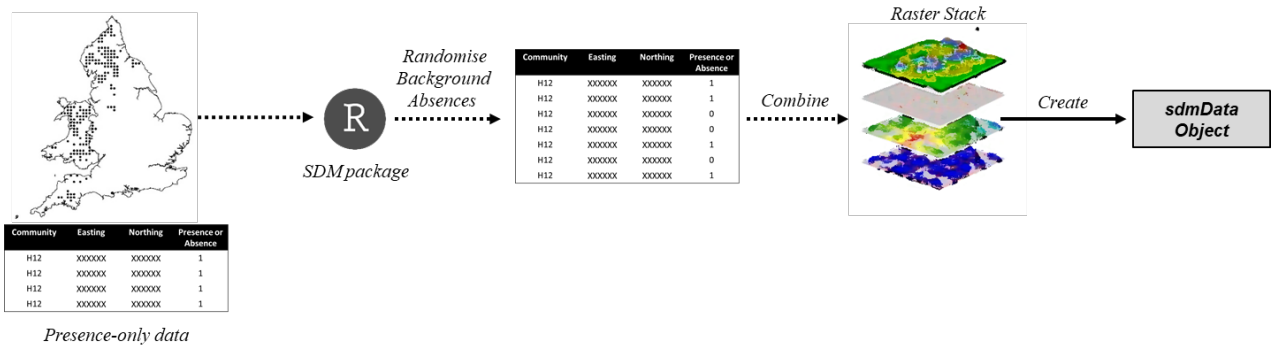


Fig. 4. 3 Step 1: The first steps of generating a sdmData object that contains vegetation community data and predictor information.

A formula interface is used to define an SDM model using the 'sdm' function (Naimi and Araújo, 2016) as 'community ~ explanatory variables'. SDMs were fit using five methods, i) generalised linear models (GLM), ii) random forest models (RF), iii) support-vector machine models (SVM), iv) maximum entropy (MaxEnt), v) maximum likelihood (MaxLike). Bootstrapping replication method was used for each model type, with each bootstrapping replication repeated 10 times. 'Bootstrapping' is a re-sampling method used quantify the uncertainty associated with a statistical or machine learning method and helps avoid overfitting (Pan, 1999; Abney, 2002; Borra and Di Ciaccio, 2010). The 'sdm' command returns an 'sdm results' object in R (Fig. 4.4), which was used to predict distribution and probability of occurrence of vegetation communities across England and Wales (Fig. 4.5).

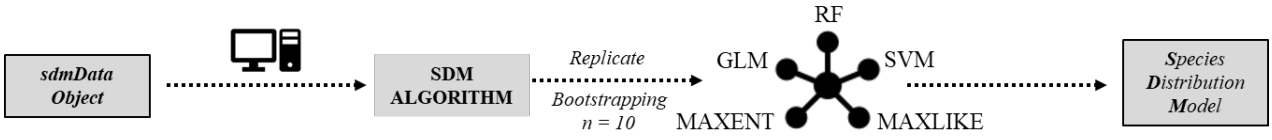


Fig. 4. 4 Step 2: Using the sdmData object generated in step 1 to run SDM algorithms for 5 model fitting methods to generate the SDM.

A raster map for the five prediction methods was generated as a raster stack for each vegetation community. The trained model was run using the different models generated by bootstrapping re-sampling as additional 'testing' data.

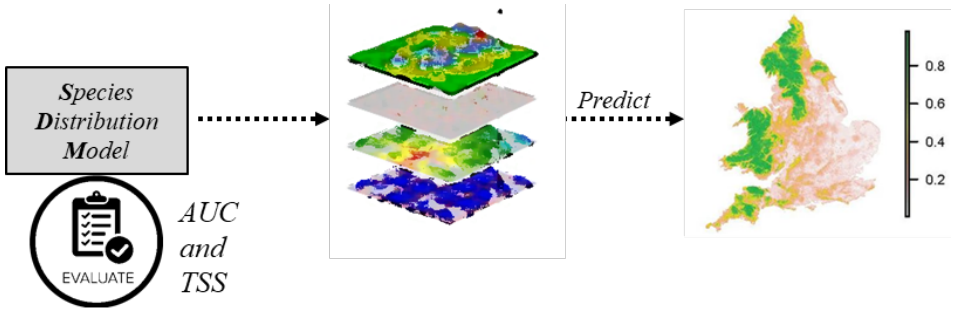


Fig. 4. 5 Step 3: Evaluating the SDM models and predicting the vegetation communities at National Scale.

Every SDM generated was evaluated then evaluated using the 'getEvaluation' function in the 'sdm' package (Naimi and Araujo, 2019). Two model evaluators were used, i) area under curve (AUC) and ii) true skill statistic (TSS; Franklin, 2010). The threshold for TSS was set to 0.5, the default and recommended threshold (Allouche et al., 2006; Franklin, 2010). Ten evaluation results were output for each model type, and the mean evaluation calculated for AUC and TSS.

4.2.4 Compare NVC community predictions with observed NBN Atlas records in the Rede Catchment (Objective 4)

The Rede Catchment was used as a small-scale test area to compare predictions for the NVC communities with the observed distribution of the most frequent species. Vegetation species with medium to high frequency of occurrence in each set of 10 NVC communities, i.e. constancy frequency of III, IV and V (Rodwell, 1991), were obtained from the NBN Atlas. Data found on the NBN is recorded by organisation and citizens, verified by experts, and usually accessed via an interactive website on a species-by-species basis. As distribution data was needed for a large number of species, rapid downloads were made directly into R, using the new 'NBN4R' package (Raymond *et al.*, 2019) as this is a development package available from https://github.com/fozy81/NBN4R although not the usual CRAN R package repository. The distribution data was stored as an R 'sf' object and cropped to England and Wales.

The final step was to spatially crop species distributions to the catchment of the River Rede, and display them as dot-distribution maps. As the typical grid resolution is 1 km for many records in NBN Atlas this can result in multiple records or species overlapping each other, making visual assessment difficult, therefore where necessary points were slightly repelled from each other using the 'geom_jitter' function from the ggplot2 R package (Wickham, 2016) for better visualisation. An assessment was then made of the observed NBN Atlas records and the predicted distribution of each of the selected communities in which these species occur, using the best SDM modelling method identified in Section 4.2.3 above. While this method of 'verification' could have been done for the whole of England and Wales, it was decided to choose the Rede Catchment to link this part of the research to the Ashtrees study site, the data of which was extensively used in Chapters 2 and 3. Furthermore, the high number of individual species records across a large area, such as the whole of England and Wales, would have made the output maps difficult to interpret.

4.3 Results

4.3.1 Collation of predictor data for use in SDMs (Objective 1)

2015 Land cover (LCM2015)

Fig. 4.6 provides a summary of the Land Cover Map at 1km. Note: this was originally a 25m and aggregated to 1 km grid resolution for use in the SDM predictions. In general, the eastern parts of England are arable and horticultural lands, while most of the western parts of England and Wales are improved grasslands. Heather is more prominent in the northern parts of England, across the North Pennines and towards NE England. Land cover might not be detailed enough to improve predictions and might result in low predictor importance (Bradley and Fleishman, 2008). While land cover is a categorical data set and can provide issues with the large number of co-variates, such categorical have been successfully used in SDM modelling (Thuiller et al., 2004; Bradley and Fleishman, 2008)

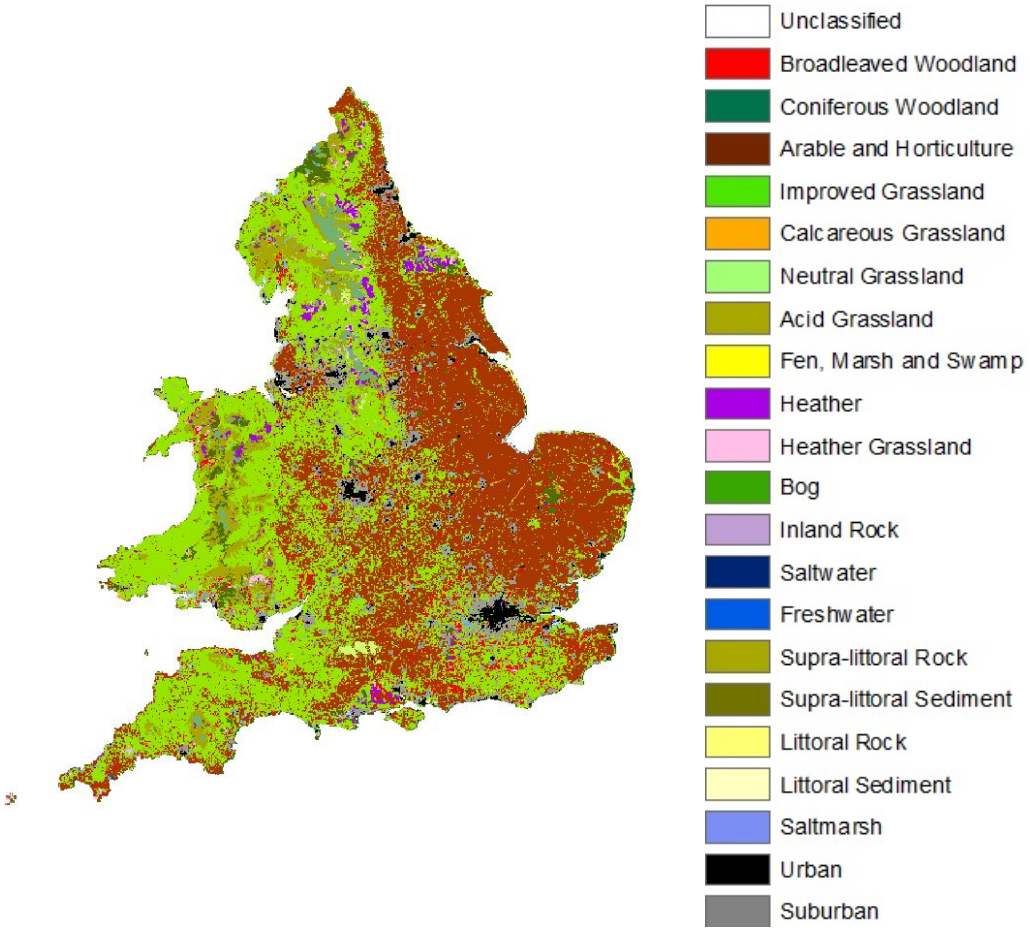


Fig. 4. 6 Land Cover 2015 (LCM2015) for England and Wales with 21 target classes. LCM2015 data was obtained from EDINA Digimap (https://digimap.edina.ac.uk/roam/download/environment) under the OS Open Data term and conditions.

Topography

Fig. 4.7a provides a summary of the digital elevation model used for the SDMs. Figs. 4.7b and Fig. 4.7c summarise the slope and aspect for the study area. Highest altitudes are generally found in Wales and Northern England, with highest altitude across the North Pennines towards the Scotland border. Highest slope gradients are in the NW of England, while aspect is variable across the country.

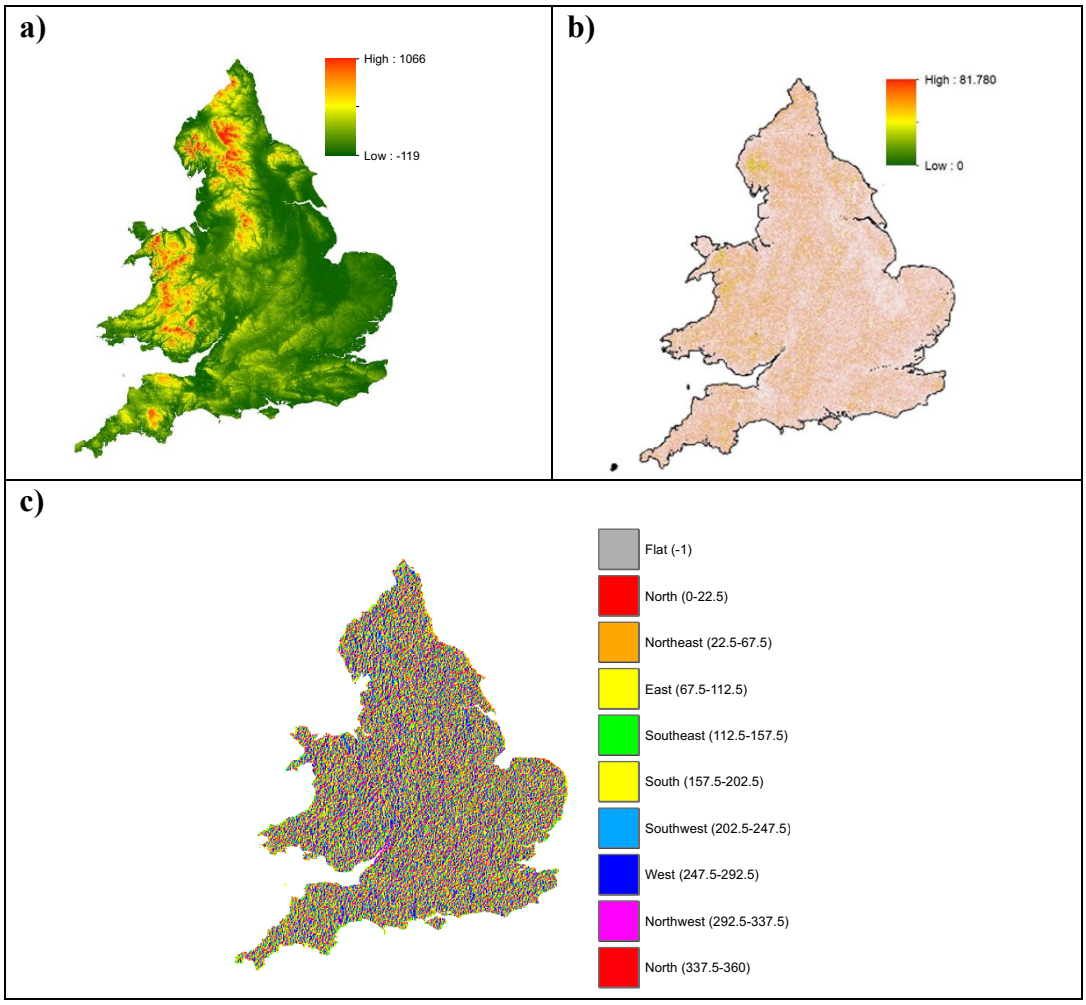


Fig. 4. 7 a) Digital Elevation Model (DEM, m) for England and Wales. ASTER Global DEM data was obtained from USGS Earth Explorer (https://earthexplorer.usgs.gov/), b) slope (º) computed for England and Wales, ranging from 0º to 82º, c) aspect (º) computed for England and Wales.

Meteorology

Temperature and rainfall (1988-1992) are summarised in Figs 4.5a-4.5e. In general, temperature is always lower in most inland surfaces (and higher altitudes) in Wales and in the north England, ranging from the North Pennines to the Scotland border. The eastern parts

England experience much higher temperatures than England and Wales. The western parts of England and Wales experience higher rainfall than the rest of the UK, becoming drier towards the east.

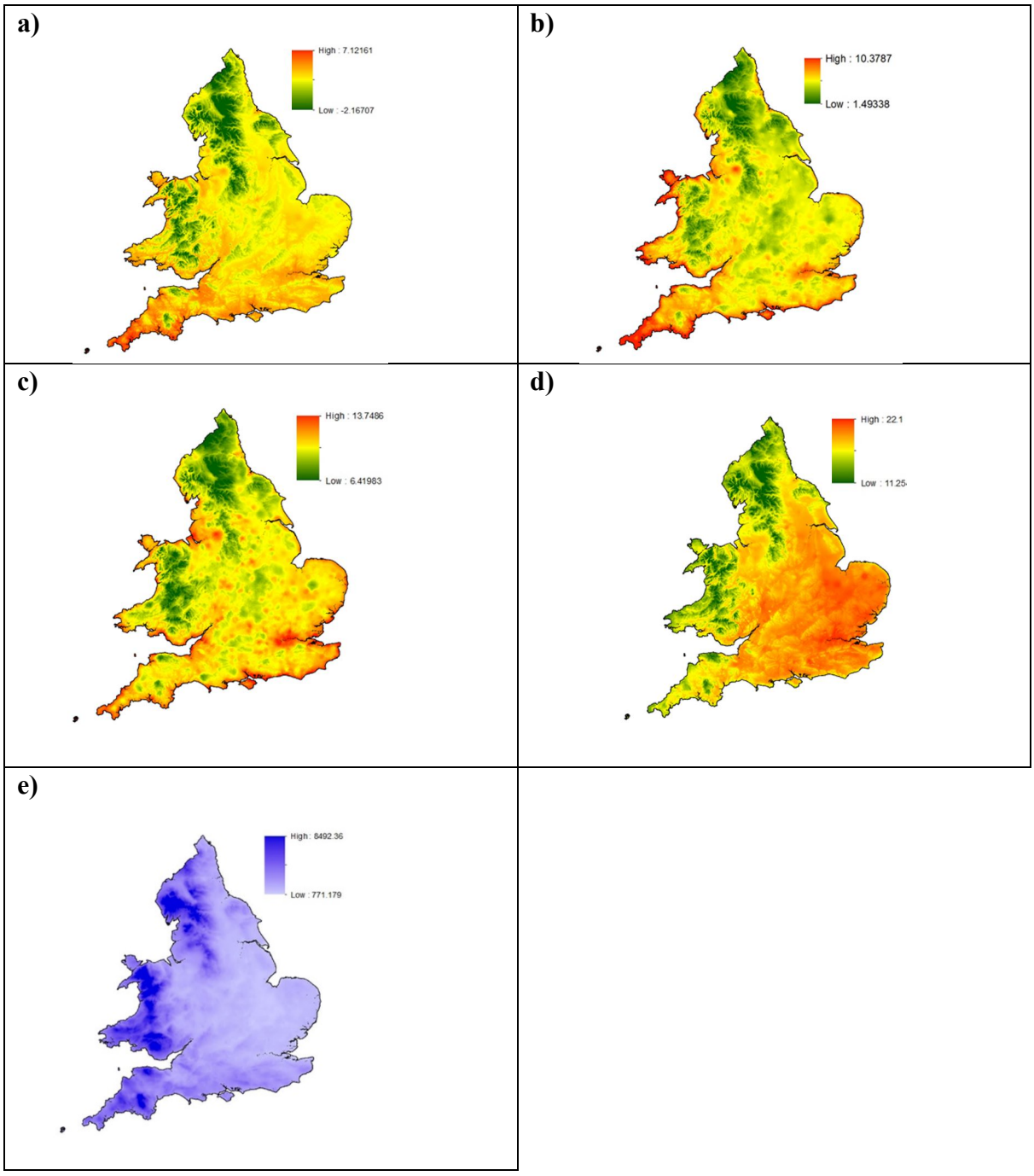


Fig. 4. 8 Mean temperatures (ºC) and rainfall between years 1988 and 1992. Data obtained from Met Office, 2018, a) winter minimum, b) winter maximum, c) summer minimum, d) summer maximum and e) rainfall.

Soil data (NATMAP)

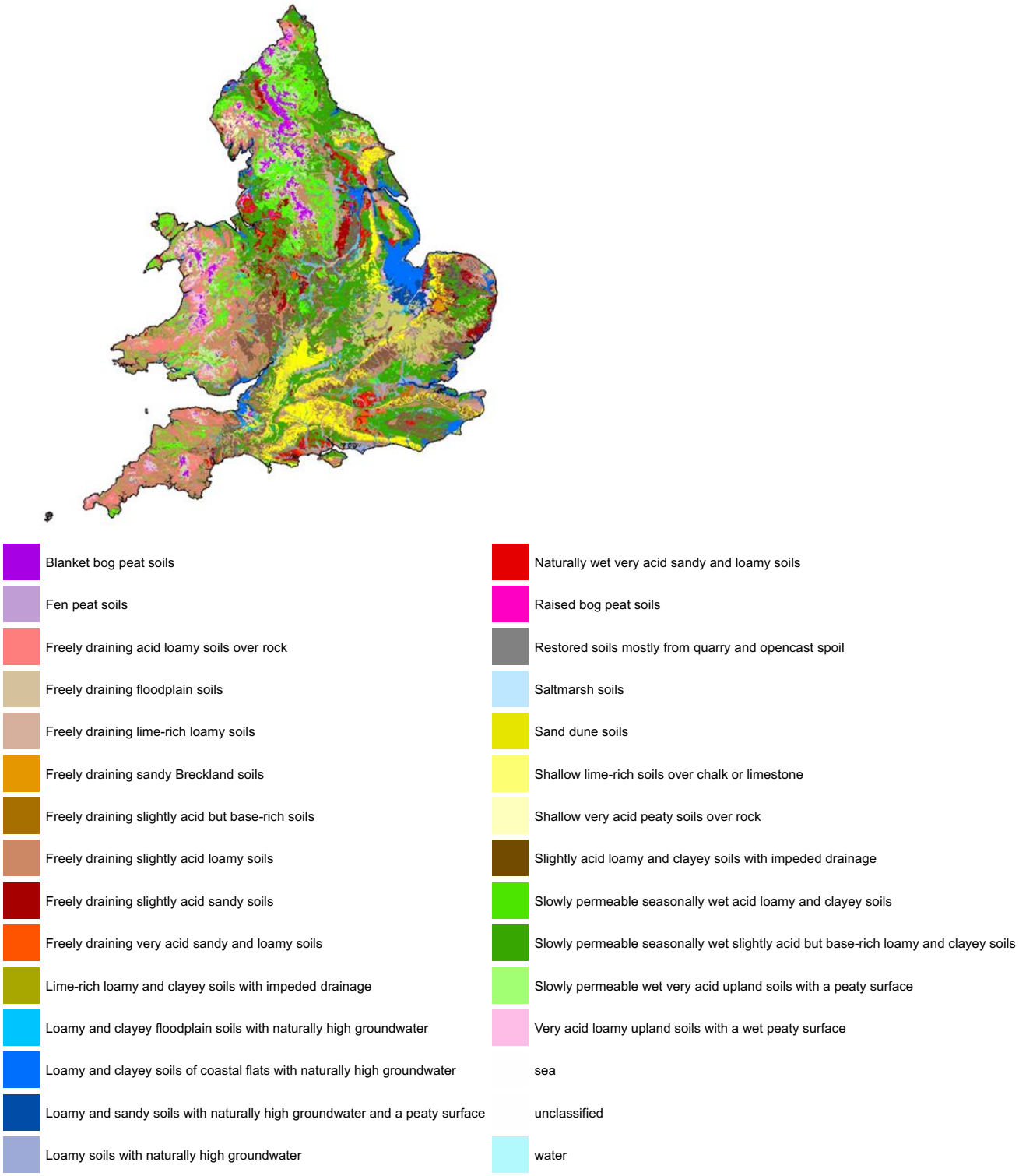


Fig. 4. 9 Soil map for England and Wales. Soil data was obtained from Cranfield University (2017) The National Soil Map and Soil Classification.

Soil map for England and Wales is provided in Fig. 4.9. Most of Wales has acid loamy soils with some blanked peat bogs, while across the North Pennines to the Scottish border soils become acidic and mixed between clayey and loamy. Blanket peat bogs are also observed. Towards the south of England, soils become more lime-rich or sandy and loamy. Similar to land cover, soil data provides the challenge of being a categorical environmental predictor and expected to have lower predictor importance.

Landsat Enhanced Vegetation Index

The EVI converted satellite data are shown in Fig. 410. Some caution is needed in the use of this map, as there appeared to be an artefact in the satellite data to the north west of the map, from NE England to mid-Wales. Areas in grey are related to urban and suburban areas (main cities) while red areas seem to be artefacts in the satellite data, showing very low EVI values (< -0.07).

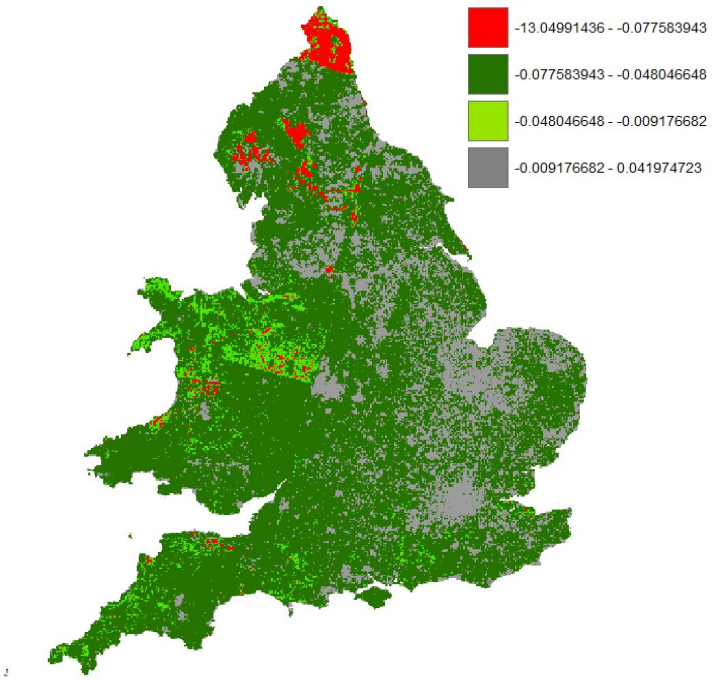


Fig. 4. 10 LANDSAT 8 EVI satellite data for England and Wales.
Data obtained from USGS, 2019; https://earthexplorer.usgs.gov/.

4.3.2 Select target NVC communities and collate distribution map data (Objective 2)

The list of NVC upland communities identified from the UK BAP and supplemented from those that were found on Ashtrees Dipper (Chapter 3) is shown in Table 4.1. 10km distribution maps were created for all these communities for input into the SDMs, with a

national total of 138 records in CG, with CG12 and CG14 being very rare communities; 401 records in H (only 1 record for H16 and H22); 1036 records in M with M26 and M38 both rare mires with only 3 and 4 records respectively; 147 records in U where 13 only has 2 records; 137 records in W classification. The communities occurring at Ashtrees are very common across England and Wales with a maximum of 270 records for U4 community. U2 has the lowest records (63). Ten of the UK BAP communities were selected at random and these are indicated in bold in Table 4.1.

Table 4. 1 NVC codes for vegetation communities that are of interest in this chapter and predicted for England and Wales. (The number of records for each community is provided in brackets).

UK BAP					**Ashtrees**
Calcicolous grasslands (CG)	**Heaths(H)**	**Mires (M)**	**Calcifugous and montane (U)**	**Woodland and scrub (W)**	**Communities found at Ashtrees**
CG9 (37)	H8 (93)	M1 (18)	U7 (14)	W9 (124)	**H12 (184)**
CG10 (90)	**H10 (82)**	M2 (71)	U10 (16)	W19 (13)	**H9 (106)**
CG11 (8)	H13 (4)	M3 (52)	U13 (2)		**M15 (116)**
CG12 (1)	H14 (4)	**M4 (89)**	U15 (5)		**M19 (139)**
CG14 (2)	H15 (2)	M5 (22)	U16 (49)		**MG10 (157)**
	H16 (1)	M8 (7)	U17 (28)		**MG6 (166)**
	H18 (148)	M9 (23)	**U21 (33)**		**U2 (63)**
	H19 (12)	M11 (17)			**U4 (270)**
	H21 (54)	**M16 (75)**			**U5 (212)**
	H22 (1)	M17 (39)			**U6 (145)**
		M18 (82)			
		M20 (141)			
		M21 (57)			
		M25 (197)			
		M26 (3)			
		M27 (50)			
		M29 (48)			
		M37 (41)			
		M38 (4)			

4.3.3 Development and testing of NVC species distribution models (Objective 3)

4.3.3.1 Overall model accuracy and importance of individual predictors

Fig. 4.11 summarises the results each of the five types of SDMs using area under the curve (AUC) and Fig. 4.12 the true skill statistic (TSS). AUC is a threshold-independent evaluator while TSS is threshold-dependent. In model evaluation metrics, the same trend in accuracy between model types is evident, with the random forest (RF) method showing highest accuracy, followed by maximum entropy (MaxEnt). MaxLike, a modified, newer technique for species distribution modelling, shows lowest accuracy for SDMs.

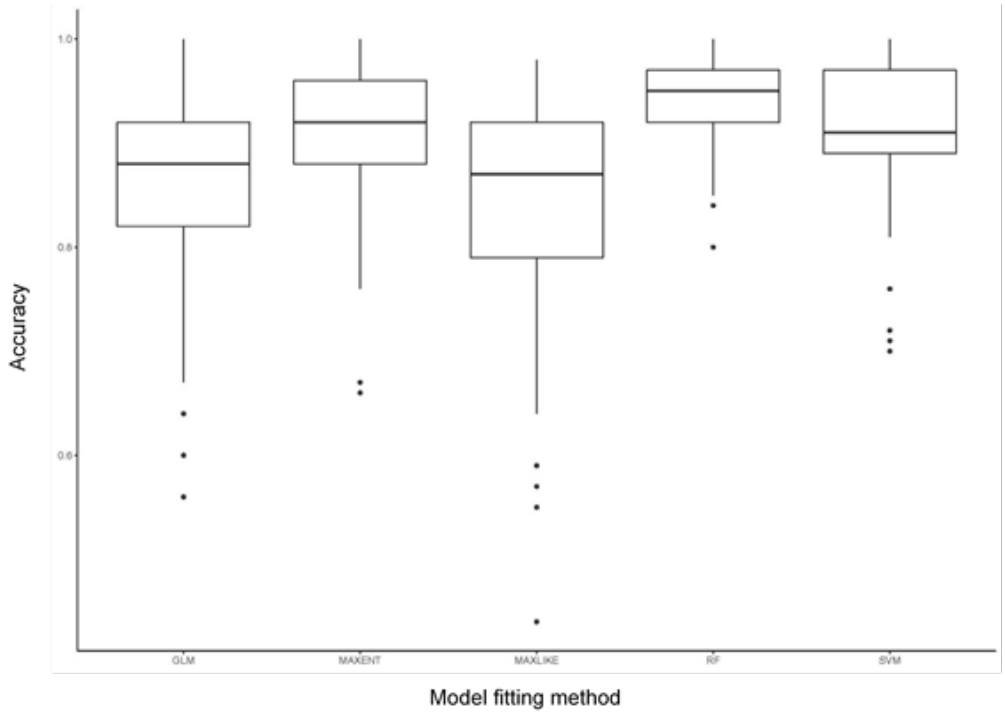


Fig. 4. 11 Mean AUC across all the SDMs generated comparing the different statistical methods used, indicating reliability of each statistical method. AUC is threshold independent.

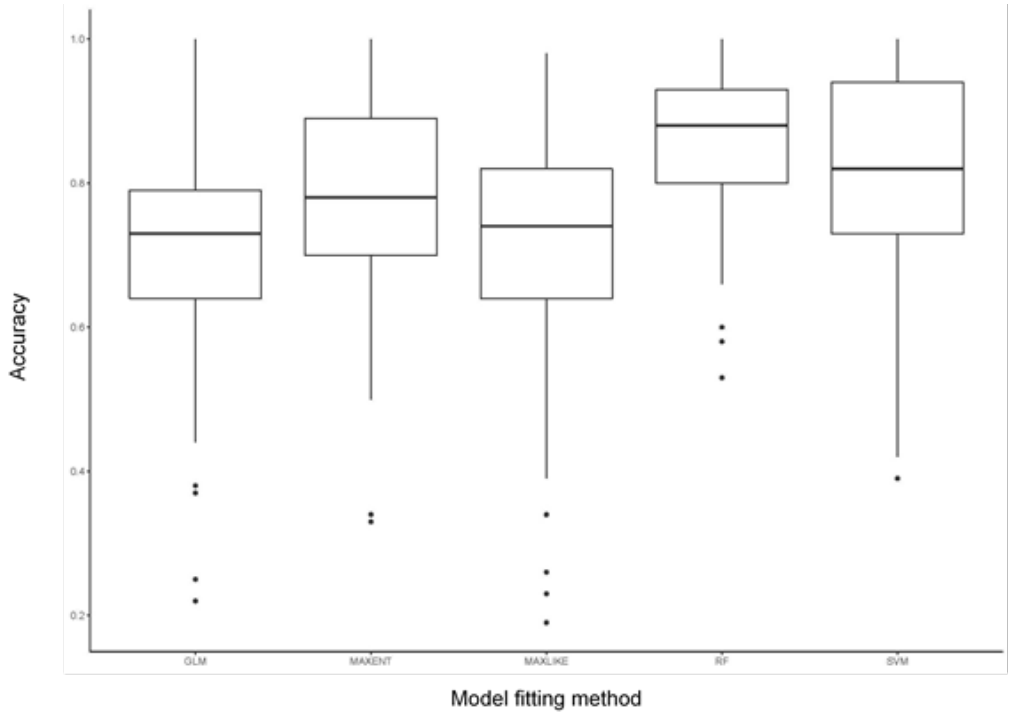


Fig. 4. 12 Mean TSS across all the SDMs generated comparing the different statistical methods used, indicating reliability of each statistical method. TSS is threshold dependent.

The overall importance of each model predictor, including topographic, meteorological, soil, satellite and land cover data, are summarised in Fig. 4.13. AUC and COR (co-linearity between predictors and SDMs) metrics both showed similar trends in importance of contribution of predictors to the models, with temperature and rainfall being the most important predictors that SDMs use for NVC vegetation community distribution modelling. Minimum temperatures in summer (TminS) and winter (TminW) had highest importance ($TminS_{AUC}$ = 0.3, $TminS_{COR}$ = 0.47, $TminW_{AUC}$ = 0.35, $TminW_{COR}$ = 0.46) by maximum temperature for summer ($TmaxS_{AUC}$ = 0.28, $TmaxS_{COR}$ = 0.42) and winter ($TmaxW_{AUC}$ = 0.24, $TmaxW_{COR}$ = 0.38). Land cover (LCM) appeared to have the lowest level of importance when developing SDMs as well as lowest co-linearity. COR results indicate the influence of each predictor across a linear spatial gradient (Smith and Santos, 2019). Many of the environmental layers might follow a linear gradient across the landscape, resulting in a spatially smoothed distribution that is autocorrelated (Smith and Santos, 2019). Fig. 4.13 shows that the predictors of higher importance are those that are more likely to form such linear gradients (e.g. temperature, rainfall and DEM) while low scoring predictors (landcover, slope and soil) show very little linearity across the landscape.

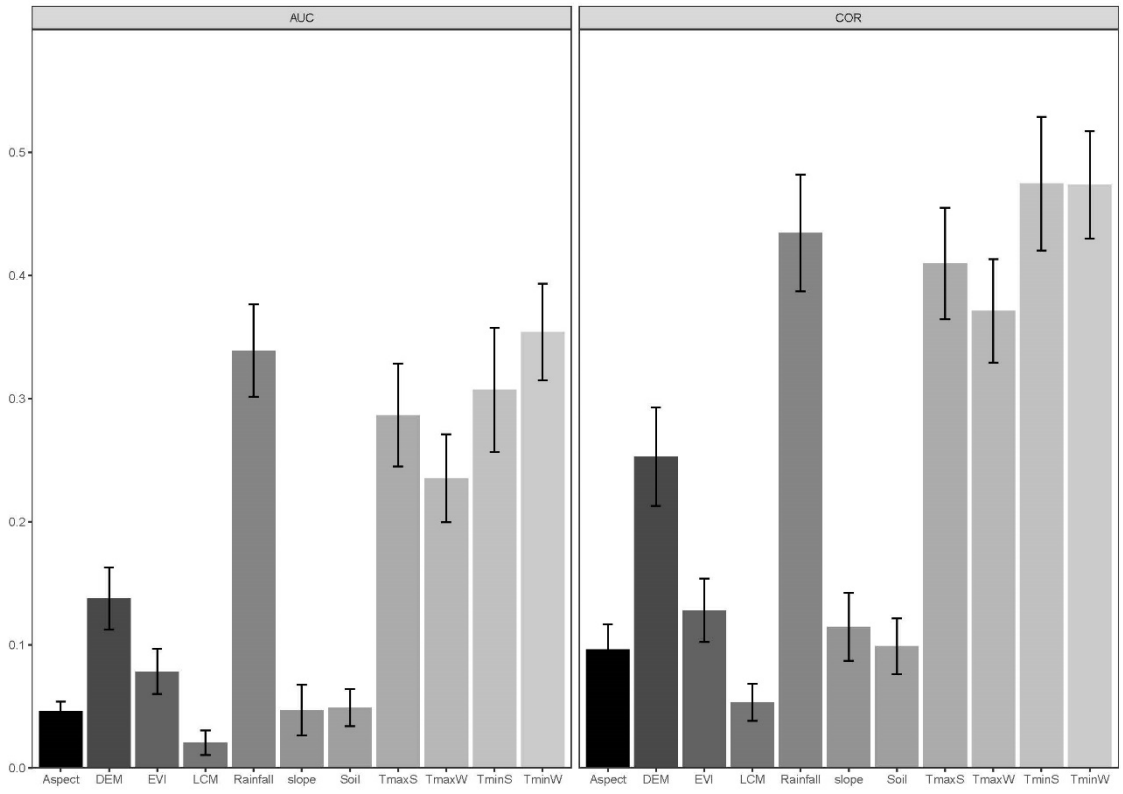


Fig. 4. 13 Predictor importance (probability score from 0 to 1) for SDM models using AUC ±S.E (left) for evaluation and COR ±S.E (right) for predictor co-linearity.

4.3.3.2 Vegetation community predictions across England and Wales

As might be expected, the patterns displayed in the predicted maps across England and Wales for the 10 Ashtrees and 10 UK BAP communities were extremely variable, and appeared to depend on the SDM model used, and whether the NVC community was widespread or locally distributed. While models were generated for each of the fourty-four vegetation communities, only results for RF model outputs for 20 selected communities are provided in Appendix 2. A qualitative visual comparison was made of the various model predictions and observed NVC distribution maps. The qualitative assessment was on a scale of 1 to 5, where 1 was judged to be a poor agreement between observations and predictions, and 5 an excellent agreement, and this is shown in Table 4.2. In general the random forest (RF) model gave the most accurate predictions, and MaxLike the poorest, which accords with the quantitative assessment of overall model performance shown earlier in Table 4.1. There was a tendency for more accurate predictions to be made at the Ashtrees communities, which may reflect the underlying rarity of the UK BAP communities.

To demonstrate the differences in model predictions and their relationships with the numbers of observed records, four communities are described in detail. H12 and U4, both with large numbers of observed records, selected from the Ashtrees list, and H10 and U17, both less frequent, and from the UK BAP list. In addition, H12 and H10 had relatively high model rankings (Table 4.2) whereas U4 and U17 had poor rankings.

Table 4. 2 Qualitative ranking of model predictions for 10 UK BAP and 10 Ashtrees vegetation communities across England and Wales on a scale of 1 to 5, with 1 = 'poorly predicted' and 5 = 'well predicted'. Green rows = overall well-predicted, red rows = overall poor predictions.

NVC	GLM	SVM	RF	MaxEnt	MaxLike
Ashtrees					
H9	5	5	5	4	1
H12	5	4	5	5	1
M15	3	5	5	4	2
M19	5	4	5	3	2
MG6	5	3	5	5	1
MG10	2	3	5	5	4
U2	4	5	5	5	1
U4	4	2	4	5	1
U5	5	4	5	5	1
U6	4	3	4	4	2
Randomly selected					
H10	5	4	5	5	3
H18	3	4	4	4	1
H19	5	1	5	4	3
M4	4	2	5	4	3
M16	3	4	5	2	2
M18	4	2	4	5	1
M21	5	3	5	4	3
M37	4	4	5	5	1
U17	4	1	5	2	2
U21	4	3	5	5	1

Both H12 (heath) and U4 (calcifugous grassland/montane) are very common communities (184 and 270 records respectively; Table 4.1) widely distributed across the Lake District, Pennines, North Yorkshire Moors, South West England and Wales (Fig. 4.14). Although the U4 community has a higher number of records than H12, its occurrence prediction was the 'poorest' of the 10 Ashtrees communities (Table 4.2). For both communities, GLM, RF and MaxEnt models provided robust outcomes while MaxLike was poor in prediction. SVM operated poorly for both the U4 and H12 communities, with predicted probabilities generally

over 0.4 or 0.5 across the whole area. Fig. 4.14 below shows the differences between SVM, RF and MaxLike model outputs. SVM slightly over-estimated the probability of occurrence of the U4 community, especially towards the south of England. This over-estimation was also observed for the H12 community, but to a lesser extent. Results for GLM and MaxEnt models were similar to RF.

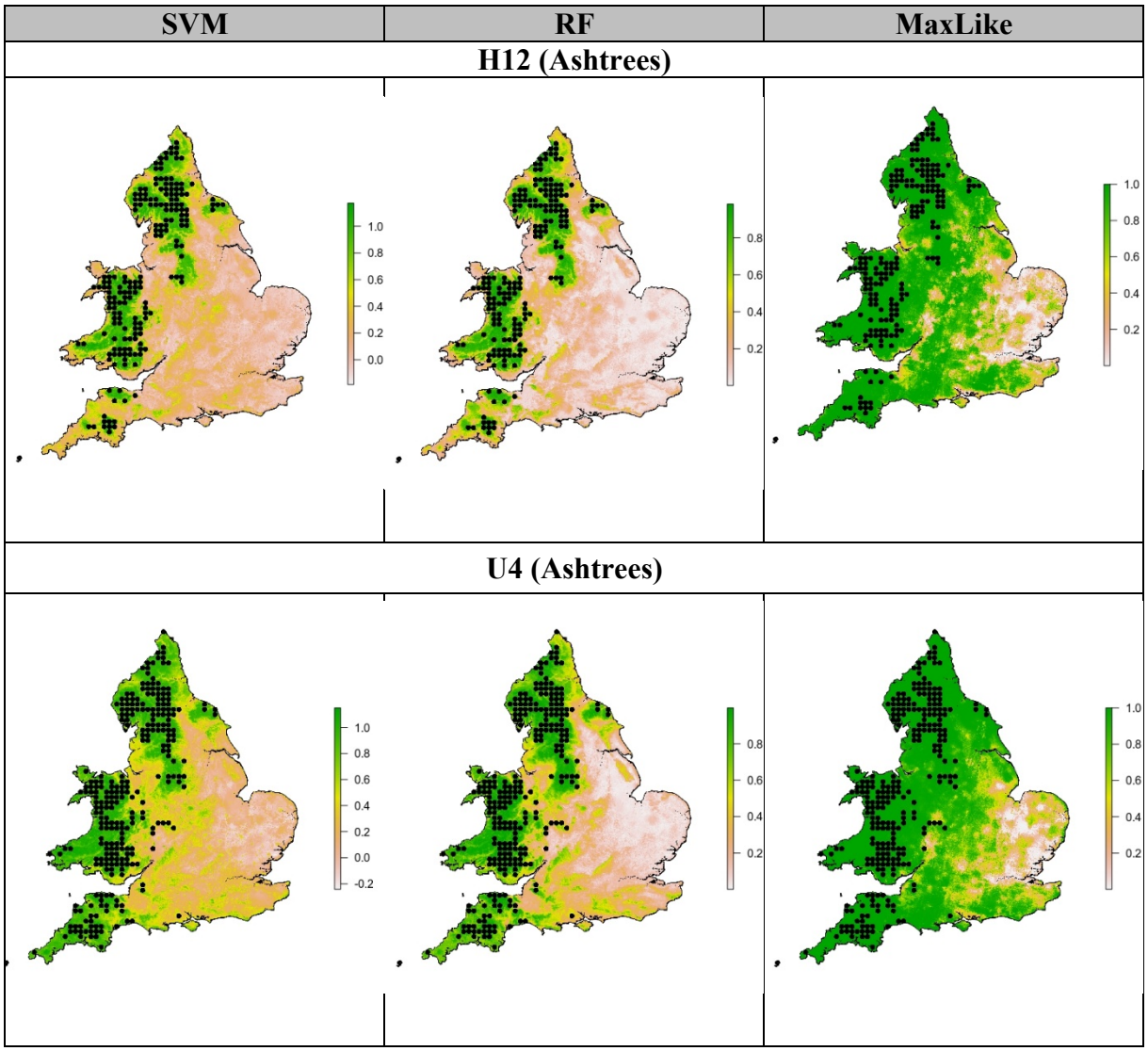


Fig. 4. 14 Comparing H12 and U4 vegetation communities from the Ashtrees list and their predicted outputs by SVM, RF and MaxLike. Black points are the recorded data of these communities.

H10 (heath) and U17 (calcifugous grassland/montane) communities randomly selected from the UK BAP are less frequent across these same upland areas of England and Wales with 82 and 28 records respectively, mainly restricted to the Lake District, North Pennines and parts of Wales (Fig. 4.15). For these less frequent communities RF produced very robust predictions of occurrence for both communities whereas SVM over-estimated probability of occurrence across the whole of England and Wales with few areas with a prediction

probability of less than 0.4 (Fig. 4.15). In contrast MaxLike overestimated occurrence in many parts of the uplands (probabilities over 0.8), but was more accurate in lowland or south eastern areas with probabilities less than 0.1. Comparing predictions for these less frequent UK BAP communities with those for the more common communities in the Ashtrees list suggests that RF is the most robust system overall, and least sensitive to the numbers of observations in the underlying data. While there is a slight tendency of such systems to provide an 'overfitted' model, the use of testing and training data sets in the original SDM development as well as random re-sampling of the data multiple model generations can decrease the possibility over such 'overfitting'.

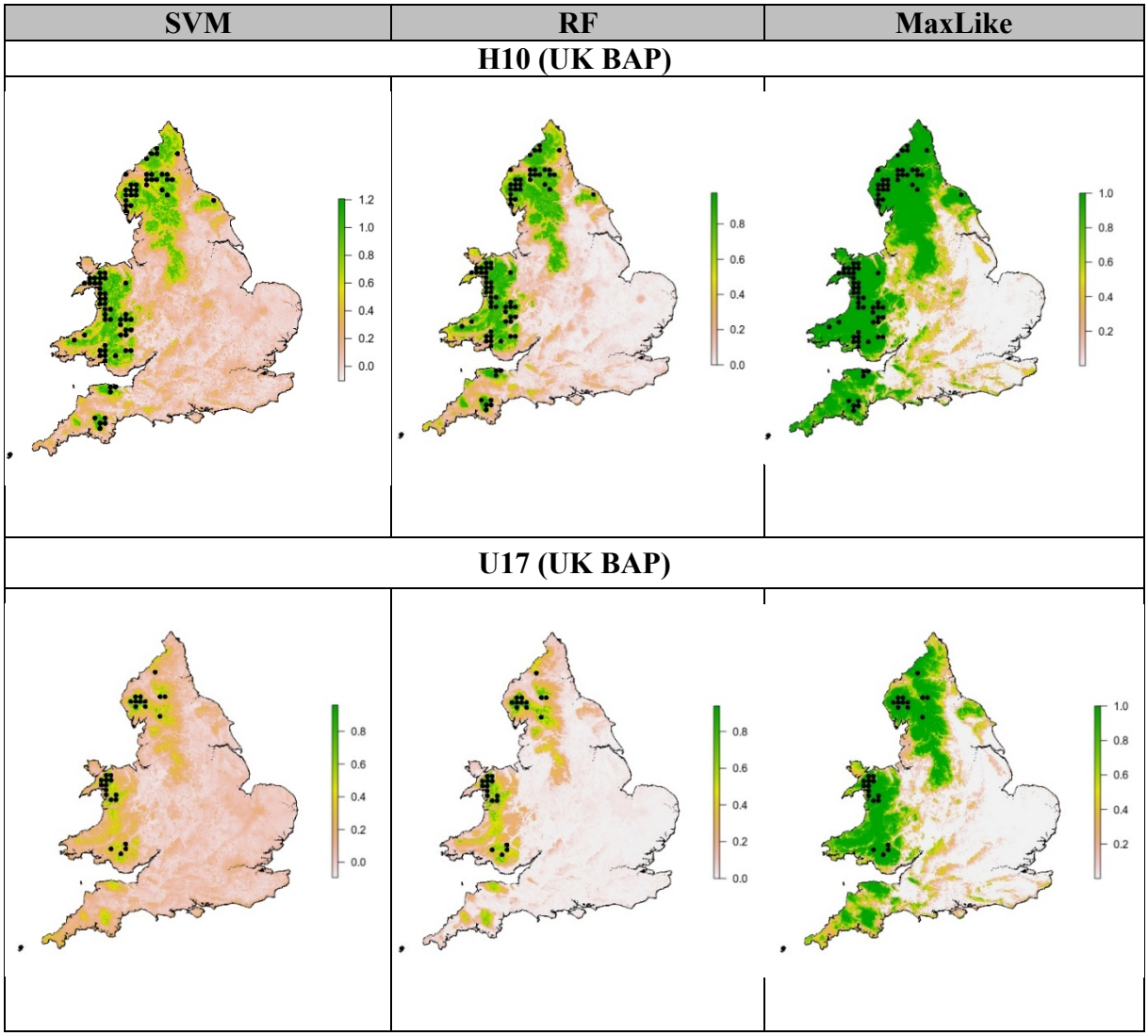


Fig. 4. 15 Comparing H10 and U17 vegetation communities from the randomly selected UK BAP list and their predicted outputs by SVM, RF and MaxLike. Black points are the recorded data of these communities.

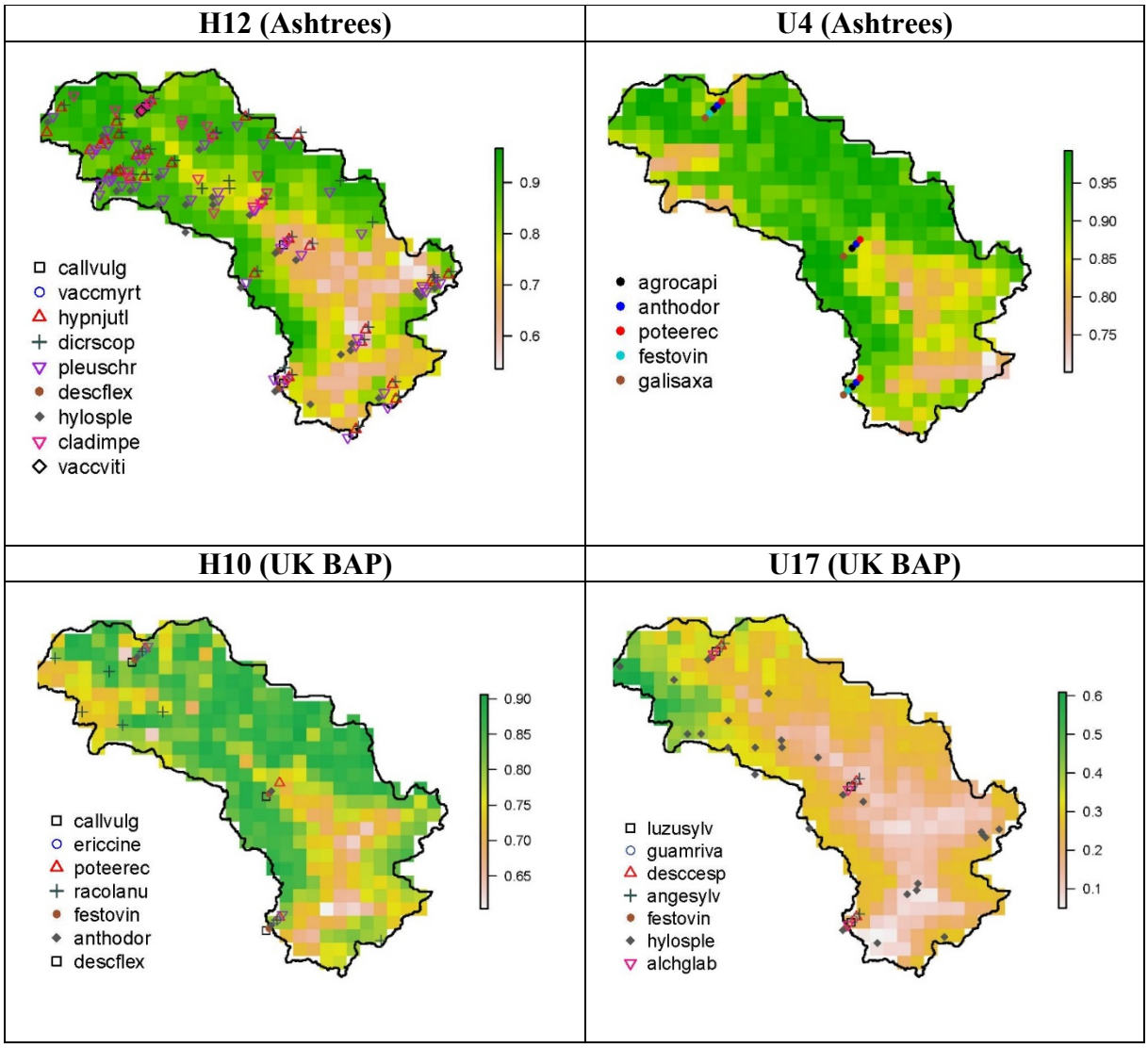
4.3.4 Compare NVC community predictions with observed NBN Atlas records in the Rede Catchment (Objective 4)

The Rede Catchment in Northumberland was used as a case-study area to map species distributions (species of constancy III, IV and V) recorded in the NBN Atlas database and determine if these fall in areas of high predicted probability of occurrence (by SDMs). Full results for both sets of 10 communities are in Appendix 2.3 and 2.4, but for consistency results for species within the same four communities described in Section 3.3 are detailed here. These are H12 and U4 (Ashtrees, good and poorer random forest SDM model predictions) and H10 and U17 (UK BAP, good and poorer SDM model predictions respectively). Random forest SDMs were identified as most robust in Section 4.3.3, therefore only these are considered here.

Table 4. 3 Random forest (RF) predicted percentage occurrence of example vegetation communities extracted from distribution of species in the NBN atlas.

	Ashtrees		**UK BAP**	
Species	**H12**	**U4**	**H10**	**U17**
Agrostis capillaris	-	89.34	-	-
Alchemilla glabra	-	-	-	25.00
Angelica sylvestris	-	-	-	22.89
Anthoxantum odoratum	-	89.69	80.49	-
Calluna vulgaris	69.37	-	79.86	-
Cladonia impexa	84.28	-	-	-
Deschampsia flexuosa	72.50	-	80.57	23.50
Dicranum scoparium	82.93	-	-	-
Erica cinerea	-	-	82.06	-
Festuca ovina	-	91.24	-	27.97
Galium saxatile	-	89.67	-	-
Geum rivale	-	-	-	22.89
Hylocomnium splendens	79.50	-	-	23.81
Hypnum jutlandicum	84.76	-	-	-
Luzula sylvatica	-	-	-	25.48
Pleurozium schreberi	82.78	-	-	-
Potentilla erecta	-	89.80	80.65	-
Racomitrium lanuginosum	-	-	80.15	-
Vaccinium myrtillus	74.43	-	-	-
Vaccinium vitis-idaea	91.42	-	-	-
Mean % probability	**80.22**	**89.95**	**80.63**	**24.50**

Community constituent species were plotted on the predicted vegetation maps as a method of assessing the reliability of predictions (Fig. 4.16), and mean percentage probabilities calculated (Table 4.3). Constancy species for both H12 and U4 had high predicted mean probabilities, with 80.2% and 90.0% respectively (Table 4.3). Note that U4 actually performed worse on the overall SDM assessment for England and Wales than H12, and care is needed in assessing these results. The NBN Atlas has over 30 species records for constancy species in H12, but only 3 records for U4. This makes comparisons between the models more difficult, and reflects the challenges in using citizen science data (see Discussion). Likewise, for H10 and U17 the number of NBN Atlas records is quite different (Fig. 4.16), but here the SDM model performance in the Rede Catchment accords with SDM model performance across England and Wales (Table 4.3, 80.6% and 24.5% respectively).



4.4 Discussion

The datasets used to generate the NVC and its published associated floristic tables provide a valuable resource, but have limitations. In particular, the NVC maps are only available at 10km resolution, and whilst the Rodwell (1988) maps have been updated, the resolution is still coarse. Bradter *et al.* (2011) also used random forest approaches to predict NVC distributions, at fine spatial resolution (5m), in combination with air photography information, but for a relatively small geographic area (Yorkshire Dales National Park). In contrast, in this thesis part of the aim was to make predictions at both national (England and Wales) as well as local (Rede Catchment) scales, and evaluate several different modelling approaches.

4.4.1 Collation of predictor data for use in SDMs (Objective 1)

Vegetation communities depend on both environment and management (Ewald, 2003), although the latter is difficult to quantify at individual sites. Remote sensing provides numerous way of obtaining freely-available environmental data at very high resolution and across multiple spatiotemporal scales (Lausch *et al.*, 2013). Meteorological data (temperature and rainfall) and topographic data (DEM, slope and aspect) were deemed as important predictors because of their ability to show differences between the northern, southern, western and eastern parts of England and Wales. Climate and topology can influence the distribution

Fig. 4. 16 Comparison of probability of occurrence of H12 and U4 Ashtrees communities and H10 and U17 UK BAP communities using species data from the NBN Atlas to assess reliability of community predictions.

and growth of different community types. Often a distinction is made between direct and indirect environmental variables in SDMs. Variables that have a direct influence on vegetation physiology and growth include temperature, rainfall and soil type (Austin, 2002; Franklin, 2010). Indirect variables include elevation, slope and aspect; these may affect seed dispersal etc. (Austin, 2002; Austin, 2005; Franklin, 2010; Bradter *et al.*, 2011).

In this study the indirect predictor elevation showed moderate to high level of importance in predicting vegetation communities across England and Wales. Elevation shows large differences across the England and Wales. This elevation cline is important for ecological diversity and distribution of species (Austin and Smith, 1990; Lomolino, 2001; Austin, 2002; Barve *et al.*, 2011; Hof *et al.*, 2012). Both aspect and slope are measures of exposure of vegetation to solar radiation (Phillips *et al.*, 2004; Bennie *et al.*, 2006; Bradter *et al.*, 2011).

Slope and angle are important features in determining phytosociological patterning and distribution at multiple ecological scales (Bennie *et al.*, 2008). This is because vegetation types can change composition across small distances with changes in slope and aspect (Billings, 1952; van der Maarel, 1996; Stage and Salas, 2007; Bennie *et al.*, 2008; Hernandez *et al.*, 2008; Måren *et al.*, 2015).

Remote sensed data does not fall neatly into either the direct or indirect predictor categories, as it depends on the vegetation. The power of remote sensed data is, however, increased when used in combination with other environmental data sources such as meteorology, altitude etc. (Austin, 2002; Rushton *et al.*2004; Elith *et al.*, 2006). The management and data processing multi-spectral satellite data is particularly complex and CPU-intensive, but has proved invaluable in species distribution modeling and prediction (Zhang *et al.*, 2003; Araújo and Luoto, 2007; Zhang *et al.*, 2017). Multi-spectral EVI uses spectral reflectance in the near-infrared, red and blue bands to generate a composite at high resolution with strong correlation to the typical green coloration of vegetation at different growth stages (Gao *et al.*, 2009; Ke *et al.*, 2015). EVI is also a good indicator of the phenology of land cover types (Zhang *et al.*, 2003; Gao *et al.*, 2009), making it an important contribution for high-quality SDMs. The use of satellite data can thus create models that distinguish between ground objects and vegetation phenology (Zhang *et al.*, 2017). The ability of EVI to distinguish between different vegetation types means that communities can be evaluated at pixel level if the spatial resolution is sufficiently fine (Zhang *et al.*, 2003; Zhang *et al.*, 2017; Wang *et al.*, 2017). This may explain why the SDM models in this thesis gave higher importance for EVI than indirect variables such as slope and aspect (Zhang *et al.*, 2003; Gao *et al.*, 2009; Zhang *et al.,* 2017).

4.4.2 Select target NVC communities and collate distribution map data (objective 2)

Different land cover types are managed differently, depending on their conservation status as well as their agricultural importance (Thompson *et al.*, 1995; Tscharntke *et al.*, 2005; Miller and Hobbs, 2007). Different management techniques are used in different habitats across England and Wales, and while environmental management could not be taken into account as a variable, the conservation and management practices are an important component of the UK BAP. The selection of target NVCs then becomes an important tool for choosing habitats that include vegetation communities that are of conservation value within different policy legislation (Tscharntke *et al.*, 2005; Pywell *et al.*, 2011).

Target NVCs for distribution modelling were those that are under conservation status in the uplands. This is because of the high-intensity management that is applied in some areas but not in others but still have the same vegetation communities (Tscharntke *et al.*, 2005; Pywell *et al.*, 2011). The NVCs chosen also range in the number of records within a 10km grid, with some vegetation communities being very rare in England and Wales (e.g. calcareous grasslands: CG11, CG12; heaths: H14, H15, H16; mires: M8, M26, M38 and calcifugous grasslands: U13, U15) while some are comparatively common (e.g. calcareous grasslands: CG10, heaths: H8, H10 and 18, mires: M2, M4, M25 and calcifugous grasslands: U16) but are still of conservation concern because of birds (grouse, curlews etc.) and herbivore grazers (Dodd *et al.*, 1994; Thompson *et al*, 1995; (DEFRA, 2007; JNCC, 2016).

4.4.3 Development and testing of NVC species distribution models (Objective 3)

Species distribution models are primarily tools to predict distributions mostly attributed to the growing need for geographic understanding of biodiversity (Franklin, 2010; De Marco and Nóbrega, 2018). The availability of different SDM techniques has become popular in ecological modelling and conservation biology because SDMs help scientists understand biodiversity changes of high-risk areas (Franklin, 2010). SDMs also provide a holistic understanding between species distribution and their relationship with the environment (Franklin, 2005; Guisan and Zimmermann, 2000; Guisan *et al.*, 2006; Franklin, 2010; Warton *et al.*, 2015). SDMs have rarely been used to predicting vegetation communities (Bradter *et al.*, 2011). This is because community-level data is difficult to assess because of the large number of individual species that comprise each community and each of these constituent species will have its own environmental tolerances (Franklin, 2010).

This research tested 5 methods for SDM generation: GLMs, SVM, RF, MaxEnt and MaxLike. These methods have differences in their performance and application to different response variables (Franklin, 2010). Random forests have been accepted by scientists to be very reliable in fitting models of varying data set sizes and robust against over-fitting models (Kampichler *et al.*, 2010; Evans *et al.*, 2011; Belgiu and Drăguţ, 2016). Overall results (Section 4.3.3) demonstrated that RF models had highest accuracy while MaxLike gave the lowest accuracy across all models. This might be because MaxLike is less effective on

presence-background (used here) than presence-only data (Franklin, 2010; Fitzpatrick *et al.*, 2013).

However, it is better in datasets where there are only a small number of records (Royle *et al.*, 2012; Fitzpatrick *et al.*, 2013), and was better than MaxEnt for infrequent communities (Table 4.2). AUC and TSS model evaluation methods all scores highest for random forest predictions, followed by MaxEnt, SVM and GLM. Random forests build numerous independent hierarchical trees and averages the final output. Each tree is built on a sub-sample of training data and uses the predictor variables to increase the resulting tree to maximum size (Franklin, 2010; Kampilcher, 2010; Evans *et al.*, 2011; Belgiu and Drăguţ, 2016). RF methods are very good at analysing and coping with non-linearity of variables and can discriminate between classification trees of most importance and relevance (Evans *et al.*, 2011; Belgiu and Drăguţ, 2016). Both random forests and SVMs are sometimes considered as machine learning methods. Both methods are computationally intensive and use decision trees in their processing, but SVMs assume that variable data are linearly correlated and separated in space (Mountrakis *et al.*, 2011; Pouteau *et al.*, 2012), whereas RF are not constrained by this assumption. It might be that this greater flexibility of RF methods (Belgiu and Drăguţ, 2016; Franklin, 2010) that explains the better overall model predictions of NVCs in England and Wales obtained in this thesis.

4.4.4 Compare NVC community predictions with observed NBN Atlas records in the Rede Catchment (Objective 4)

The vegetation communities identified in the UK BAP were screened for their presence in the NE England. These communities together with the communities identified in Chapter 3, allowed this research to further assess the reliability and robustness of the SDMs and predictions, using the Rede Catchment as a case study. The two sets of data (UK BAP and Ashtrees) also allowed comparison of vegetation communities that are both common and rare in NE England to identify any differences in SDM and their possible dependence on the rarity of the community for robust predictions. The NBN Atlas was used in this research as a form of qualitative assessment on the robustness of predicted community vegetation data. Literature emphasised the need for pinpointing where vegetation is most likely to occur, to reduce monetary and human effort for sampling vegetation of interest and conservation value (Watt, 1947; van der Maarel, 1996; Parkes *et al.*, 2003; Fischer and Lindenmayer, 2007; Pescott *et al.*, 2015; Pescott *et al.*, 2019). H12, U4 and H10 vegetation communities resulted

in overall high predicted probability when using species presence as a qualitative (or semi-quantitative assessor). However, the U17 community, a less-frequent community across England and Wales, resulted in an overall results of 25% probability, where species of high constancy in this community were predicted across Rede Catchment. This suggest that the model can either be a poor prediction of the community or, more likely, the constituent species of this community are present in high constancy in other similar vegetation communities (e.g. *Festuca ovina* has high occurrence in H10 and U21 communities; *Hylocomium splendens* has high constancy in H12, H18 and H19 communities; Appendix 2).

There are challenges when interpreting the NBN Atlas results. First, the comparison is not an exact "like-for-like" in that the distributions of individual species were not predicted, but rather NVC communities. One theoretical approach to resolve this would be to calculate the predicted probability of every NVC in each pixel, and then weight all potential species (across all NVCs) by the probability that the NVC is present. Whilst this would produce predicted probabilities for individual species, rather than communities, the procedure would be very complex given both the number of NVCs and the large number of constitutent species. Second, the maximum and minimum predicted probabilities were quite variable in their range, for example 0.05-0.65 for U17 compared to only 0.75-1.0 for U4. This might over-inflate the apparent accuracy of U4. Finally, this effect is exacerbated if there are relatively few records in the NBN Atlas dataset for some communities, making it harder to compare observed sightings with the predictions for individual communities.

4.4.5 Citizen science and plant monitoring schemes

Using citizen science for data gathering has become an important aspect amongst scientists to engage the public and raise awareness as well potentially reducing survey costs (Silvertown, 2009; Callcutt *et al.*, 2018). The NBN Atlas is a major step forward linking citizen science to scientific research and, indirectly, monitoring schemes that are required under numerous environmental policies, national, European and international, such as the EU Habitat's Directive (Council Directive 92/43/EEC; European Commission, 1992) and EU Birds Directive (2009/147/EC; European Commission, 2009). This online database and its primary aim of engaging citizens to record species, their location and, preferably, abundance, is a big development in identifying locations of primary concern. However, the dependency on volunteers does produce gaps in the data gathered mostly because of restricted areas,

misidentification and/or low numbers of volunteers (Dickinson *et al.*, 2010; Dickinson *et al.*, 2012).

A recent vegetation monitoring scheme in the UK has been developed and launched on mapping distribution of vegetation and their abundance patterns to allow for recognised environmental change called the 'National Plant Monitoring Scheme' (NPMS; Pescott *et al.*, 2019). This monitoring scheme allows for estimation of trends in environmental conditions of high priority areas through detection of changes in vegetation patterns and community-species indicators. However, like the NBN Atlas this depends on citizen science with volunteers playing a major role in correct identification of species and use of sampling strategies (Silvertown, 2009; Tulloch *et al.*, 2013; Pescott *et al.*, 2017; Pescott *et al.*, 2019). Unlike the NBN Atlas, the NPMS takes into account abundance of vegetation species, which has a two-fold effect, i.e. ensuring vegetation communities are in healthy form and patch metrics as well as providing a solid baseline for community distribution prediction and their frequency status (Pescott *et al.*, 2017). As with any other monitoring scheme, the main limitation is identifying locations of high occurrence probability of select vegetation and their community (Franklin, 1995; Franklin, 2010; Chandler *et al.*, 2017).

This research showed that vegetation communities' occurrence and their spatial distribution can be reliably predicted. Results from this research and the comparison with species data from the NBN Atlas show that GLM, RF and MaxEnt modelling methods can predict vegetation communities of varying number of recordings accurately. Care should be taken when using SVMs and MaxLike due to the increased chances of underestimating (SVM) or overestimating (MaxLike) vegetation community distribution and probability of occurrence when data is variable.

4.5 Conclusion

This chapter investigated whether vegetation communities, from the NVC, could be reliably predicted across England and Wales. Prediction of vegetation communities at such a large scale require high-resolution topographical, meteorological and remote sensed data for reliable outcomes community distribution. The collection and collation of environmental and remote-sensed data is an essential step in developing robust species distribution models. While species distribution models are generally used to predict species-level distributions, this chapter uses the same methods to predict vegetation community-level distribution across England and Wales. The results indicated that even with NVC vegetation data is extremely variable in number of surveyed points per community, the vegetation communities could nevertheless be reliably predicted. Results demonstrated that random forest models SDM techniques had highest success of producing robust predictions, which accords with previous studies. There are, however, a number of outstanding problems that still need to be overcome in order to predict the distributions of the individual plant species that collectively constitute each NVC community.

The resulting community distribution maps of threatened and/or prioritised vegetation communities provides useful basis to identify locations where additional surveys might be prioritised for both professional researchers and citizen scientists. This would efficiently fill in gaps in records of both the NBN Atlas and the NPMS. In addition, the methods applied in this chapter can be generalised and used at any spatial scale.

Chapter 5. General Discussion

5.1 Overview

Phytosociological research plays a very important part in understanding vegetation distribution especially in areas where vegetation change is relatively stable and changes only occur over decades. Such areas include upland, montane environment and bogs. Some of these habitats fall within Areas of Outstanding Natural Beauty, Environmentally Sensitive Areas, Specially Protected Areas and Special Areas of Conservation (SACs), and are often used in agricultural practices such as sheep-grazing and are thus semi-natural environments. Survey data at different spatial scales is a very important resource to enable scientists to monitor, protect and conserve such environments. In addition, policy frameworks exist to make sure that such areas are protected with the main aim of helping habitats and communities to regenerate by ensuring suitable habitat-environmental feedback systems as well as reducing anthropogenic influence. However, policy-makers depend on scientific data to identify areas that need to be protected. Such data needs to be reliable, robust and at at the appropriate spatial scales. Obtaining this data can be difficult due to expenses, human effort as well as environmental restrictions for experts to enter certain areas.

Vegetation communities have been studied for more than 50 years (e.g. Braun-Blanquet, 1932; Watt, 1947). More recently a large number of advanced numerical techniques have been developed for multivariate analyses of *quadrat* x *species* matrices, and their relationships with the environment. These include classification methods such as TWINSPAN, unconstrained DCA ordination, constrained canonical correspondence analysis (Ter Braak, 1987), partial CCA (Legendre and Gallagher, 2001; Sanderson *et al.*, 1995) and multivariate GLM (Warton *et al.*, 2012). National systems to classify vegetation communities, such as the NVC, have been developed with the aid of some of these numerical methods. However, despite these advances much of the research has focussed on individual species rather than the whole community, and it is important to understand the dynamics and concepts from a top-down approach, i.e. community-based approach (van der Maarel, 1996). This is because within a community species do not occur in isolation, but interact with each other in both time and space. In particular, there has until now been little research into small-scale patch structure of individual species within the community as a whole in terms of identifying patch metrics and the relationship of such vegetation metrics to the environment (le Roux *et al.*, 2014). The reason for this is probably because of the considerably larger resources and effort

required for sampling at sub-quadrat scale. Recent technological advances such as unmanned aerial vehicles (UAV) have the potential to improve the quality and quantity of spatial data at fine resolution (Xiaoqin *et al.*, 2015; Anderson and Gaston, 2013). This becomes more powerful with the development of strategies to integrate across spatial scales, from sub-quadrat, quadrat, field, landscape, region and nationally, accessing more traditional sources of remote sensed data such as air photography and satellite. The power of these approaches can be increased through integration with expanding citizen science databases, and the effectiveness of the volunteer surveyors improved.

5.2 Summary of key findings

5.2.1 Chapter 2: Understanding vegetation species patches at small spatial scales

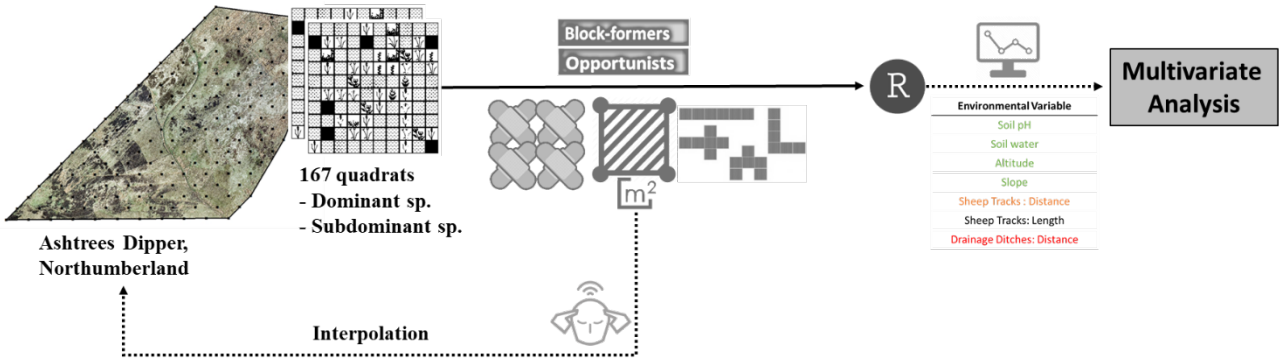


Fig. 5. 1 Graphical summary of Chapter 2 highlighting the main methods and outcomes of this chapter.

The first chapter in this thesis, summarised in Fig. 5.1, compared traditional vegetation survey data with dominant/subdominant vegetation data. Results suggested that there is strong relationship between summarised vegetation covers derived from 10cm grid data and overall visual assessments at $1m^2$ quadrat. This suggests that that upscaling to $4m^2$ quadrat data (as used in the NVC) can also be reliable.

The different types of patchiness exhibited by dominant and subdominant species correlated with soil pH, soil water, slope, aspect and herbivore grazing (using sheep tracks as an indicator of the latter). It was evident that patch area and shape were very variable across a field, even in quadrats with otherwise relatively similar environmental conditions. These differences may be a result of other, un-measured, environmental variables. Only indirect measures of herbivore grazing, trampling (Fenton, 1937; Lange, 1969) and soil drainage (Coulson *et al.*, 1990) were available and it might be that more accurate direct measures are needed. Nevertheless, RDA and 'manyglm' analysis of sub-quadrat vegetation patterns suggested that all patch metrics are influenced by soil pH, soil water content, slope and altitude (Bennie *et al.*, 2006). Dominant and subdominant patch structure (mainly number and area) are influenced by proximity to sheep tracks (Watt, 1947) and drainage ditches (Coulson *et al*, 1990; McLaughlin and Mineau, 1995), albeit to a much lesser extent than with soil pH, soil water, slope and altitude. In addition, assessment of block-forming and opportunistic species, irrespective of cover type, are also influenced by these environmental variables.

Dominant block-formers were generally found to form larger areas and very complex shapes when compared to sub-dominant block-formers. However, sub-dominant opportunists produced larger areas and more complex shapes when compared to dominant opportunistic vegetation. These may reflect changes in competitiveness between species, depending on the sizes of patches.

5.2.2 Chapter 3: Generalisable methods to classify vegetation using 'pseudo-quadrats'

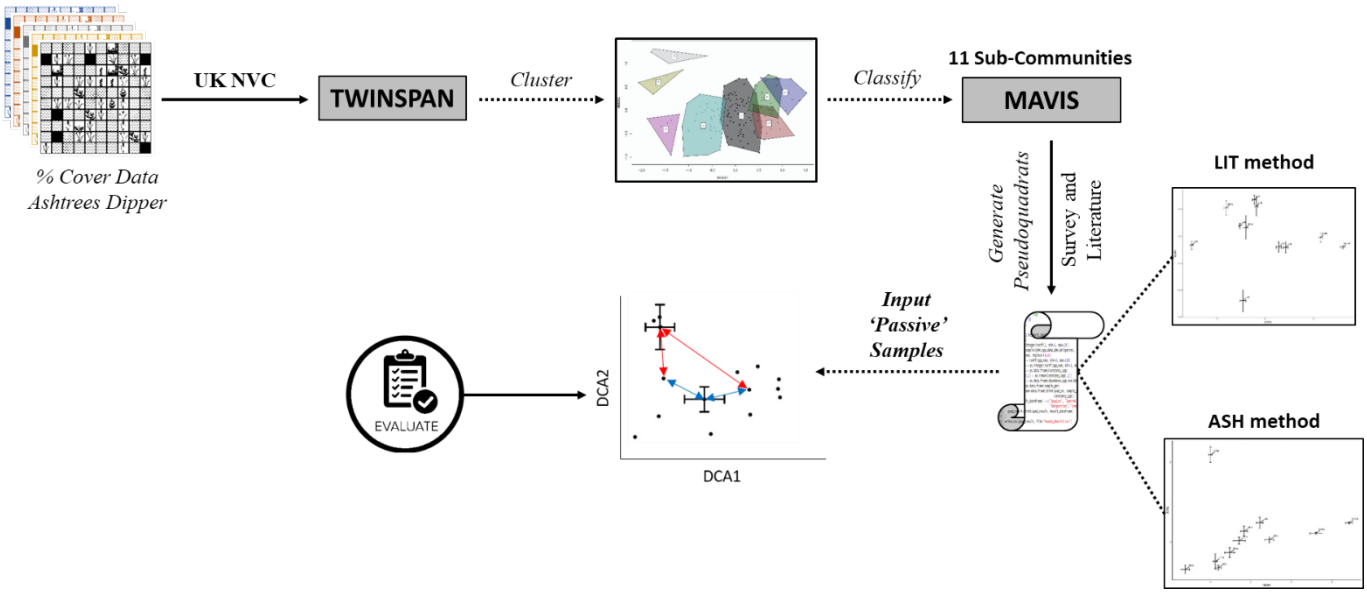


Fig. 5. 2 Graphical summary of Chapter 3 highlighting the main methods and outcomes of this chapter.

The methods developed in the second data chapter, summarised in Fig. 5.2, addressed the issues of currently available vegetation classification software using computer-generated pseudoquadrats. Pseudoquadrats provide a means to bridge data gaps, handle noisy data and aid classification of vegetation communities when sampling data is limited or was not collected via conventional methods (Butler and Sanderson, 2018). The use of pseudoquadrats provide valuable insights when classifying surveyed species into communities using traditional hierarchical and ordination techniques. Inclusion of field survey data as 'passive' samples in ordination space provides a good visual feedback to users to help them understand where their field samples sit within the context of the extant target communities, over and above a simple list of probabilities. Furthermore, previous methods to allocate field quadrats to communities, such as MATCH, TABLEFIT and MAVIS (Malloch, 1998; Hill, 1989; Hill, 1993; Smart *et al.*, 2016) have pre-requisites before they can be used, such as the methods used to survey data in the field (five 2m x 2m quadrats, randomly placed in a block of visually

homogenous vegetation for the NVC), or cannot be used on quadrats collected in isolation or from transects.

5.2.3 Chapter 4: Prediction of vegetation community distribution across England and Wales

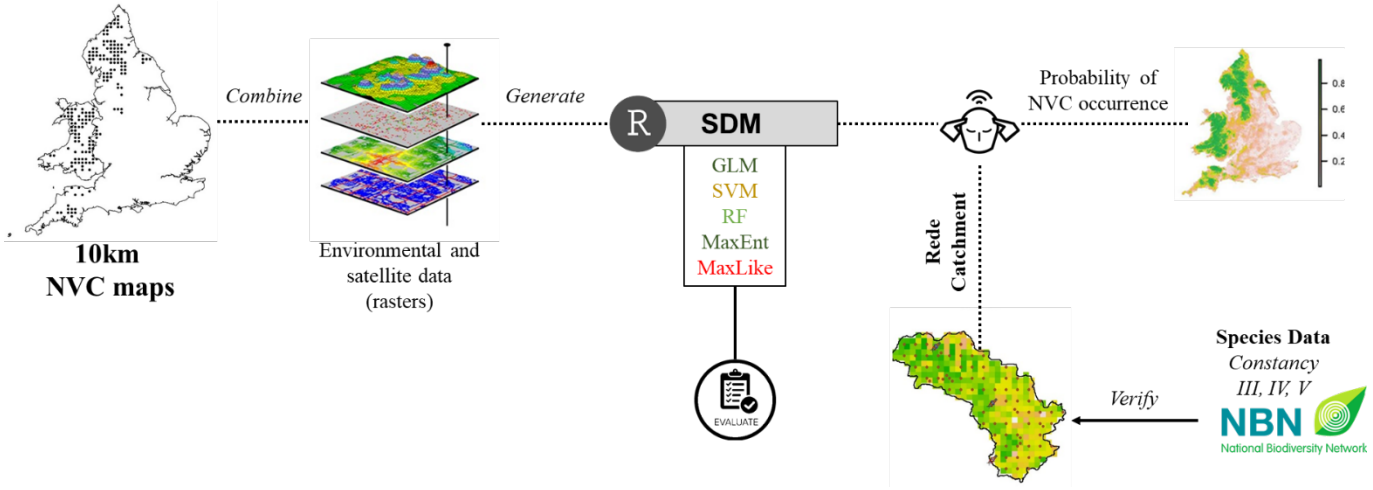


Fig. 5. 3 Graphical summary of Chapter 4 highlighting the main methods and outcomes of this chapter.

The final chapter in this thesis, summarised in Fig. 5.3, aimed to predict reliably distributions of NVC vegetation communities that are of conversation value listed in the UK BAP (JNCC and DEFRA, 2012; McLeod *et al.*, 2009). From the five different SDM modelling techniques (GLM, SVM, RF, MaxEnt and MaxLike) for England and Wales, the random forests method was the most reliable. The Rede Catchment was used as a case study to incorporate species data from the NBN Atlas to the NVC predictions at a smaller spatial scale. The NBN Atlas data is very variable between communities in terms of records for constituent species and therefore this method of assessment for predictions can be difficult. However, the resulting NVC predictions of locations for vegetation communities can help fill in the data gaps (Hampton *et al.*, 2013) in plant monitoring schemes (e.g. NBN Atlas and NPMS). These NVC predictions can reduce the risks of biases from *ad hoc* surveys, but rather target surveys to species likely to occur in specific vegetation communities even at small catchment scales (as performed here in the Rede Catchment). Identification of vegetation community hotspots also provides information for citizen scientists, allowing the public to survey particular areas and making the most of citizen science in ecology (Dickinson *et al.*, 2010; Tulloch *et al.*, 2013). Predicting NVCs that are of conservation value is important for both environmental management, stakeholders and policy makers (Sinclair *et al.*, 2010).

This is mainly because the SDM methods are simple to employ, reliable and provide invaluable insight into the locations that require protection and management.

5.3 Top-down and bottom-up approaches to vegetation community ecology

This research has been developed systematically from understanding species patch structure at sub-quadrat and quadrat scale, interpolation of species probabilities to field scale, use of computer-generated pseudoquadrats to predict communities at field scale, and finally predictions at landscape and national scales. The methodological approach from very small spatial scale to very large spatial scale means that there is also potential for a top-down approach. This means that prediction of probability of community occurrence can then use the pseudoquadrat techniques to identify species composition and their abundance within any given community. Having an approximate measure of abundance, frequency or probability of vegetation communities provides a means to predict the community constant species, and hence potentially the frequencies of these community constants. Currently, using a 'top-down' approach would still necessitate field surveys to measure patch structure at very small sub-quadrat scales. Whilst theoretically species patch structure information could be aligned with community information, in practice a large amount of additional work would be needed, probably making it impractical. Nevertheless, one route by which useful additional data might be collected at low cost would be to slightly extend existing citizen science surveys, such as NPMS, so that it includes some measures of dominant and subdominant species at small spatial scales. This data can then be used to identify trends occurring between species in different locations. The mapping of vegetation community distribution at a national scale can then allow scientists and policy makers to address issues for further conservation assessment or environmental management.

5.4 Strengths and weaknesses of the research

Each chapter in this thesis is individually-based on a particular spatial scale ranging from small 10cm and 1m scale (Chapter 2), >1ha field scale (Chapter 3) and >1km national scale (Chapter 4). The thesis as a whole tries to bridge the different spatial scales through a multi-method approach. An important concept of this thesis is that the environmental drivers (or predictors) used to understand the different processes can be substantially different at each

scale. For example in Chapter 2, site data or point-process data is required for reliable understanding of the interaction between vegetation patch patterns (assessed through patch metrics). This type of data, such as soil pH, soil water content and, to a certain extent slope, is site-specific, making the results less generalisable. However results from comparison between survey techniques (% cover abundance, dominant and subdominant) infer a reliable application to field surveys within any habitat. It is acknowledged that the relationships between the environment and the patch structures can be somewhat site-specific, and an improvement of this method would have been to utilise data from different locations of similar upland sites. This could provide a more robust analysis, and aid development of a more generalisable approach that utilises a deeper understanding of species-level patch development. This issue of site-specificity is also observed in Chapter 3 using site-specific vegetation survey data to compare methods. It should be noted that the methods used in Chapter 3 use ordination techniques, which has the advantage of providing a powerful data visualisation technique, familiar to most practising ecologists. However, the low number of vegetation communities (when compared to the whole of the NVC) imposes limits on the survey data that can be inputted in the methods described in Chapter 3. It is not practical to create a single ordination analysis that encompasses the entire NVC (R. A. Sanderson, pers. comm.), but the pseudo-quadrat approach is nevertheless sufficiently generalisable to be expanded to encompass additional communities. For example, it could be modified to a single overall classe-level analysis (i.e. aggregations of 'U', 'H', 'M', 'MG; etc.), producing one ordination plot for all the broad classes, and then lower in the hierarchy individual ordinations that contains centroids for all sub-communities within each class.

The methods described in Chapter 4 could be used to aid long-term predictions of plant community distributions, especially as a result of land use change, as well as aiding policy-makers identify potential communities or species that are potentially under threat. The data, method and results obtained in Chapter 4 provide an important 'baseline' that can be used to aid improvement in current assessment or monitoring practices at large (national) scales. While the method of using community data originally recorded at 10km has its weaknesses, it can serve as a new and important indication of where upland UK BAP, or any other particular communities are found. It is important to note that this method uses dot-distribution data from the NVC per community and it is just one of multiple data sources that could be used to map and predict community distributions.

5.5 Future work and recommendations for future researchers

Two alternative methods to predict communities and species distributions might be explored based on ideas presented in this thesis. First, community-level predictions could be created with the aid of stackable SDMs. This would involve the creation of a large number of standard SDMs for each species and the predictions for those known to occur in a given community could then be stacked. The locations where most of these species occur, especially community constants, would provide an indication of high probability that that community is actually present. Second, it might be possible to develop more sophisticated techniques to predict the distributions of individual species from the community-level probabilities presented in Chapter 4. Each NVC plant community provides a constancy score (I, II, III, IV and V) reflecting the frequency with which species have been recorded in a community. These could be converted into probabilities (0.2, 0.4, 0.6, 0.8 and 1.0) and multiplied by the probability that the community is present at that location. The most effective approach to these calculations might be to create a matrix of species by probabilities for each community (from the NVC handbooks) and multiply this with a vector of communities by probabilities, at each point in the map. Whilst the method would be numerically intensive, it might provide a more accurate method to create 1km maps of the predicted distribution of each species. The community- or species-level maps produced from either approach might have multiple uses for both researchers and policy-makers, especially with regards to species range shifts and nature conservation.

The community distribution models described in Chapter 4 could be enhanced to determine potential community distribution shifts with climate change and other such major anthropogenic influences. The use of 'hybrid' species distribution models (H-SDMs; Gavish, 2014; Singer et al., 2018) might provide substantial information on and development of models in terms of species distribution changes with climate change and major anthropogenic influences. Fig. 5.4 is a graphical summary of the development of 'hybrid' species distribution models.

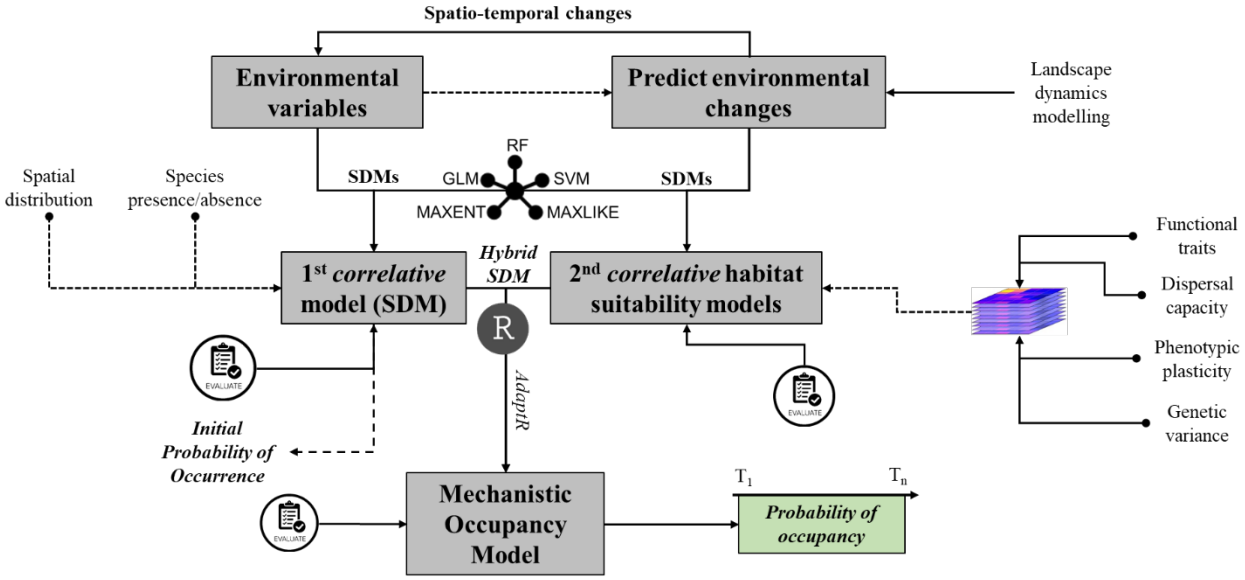


Fig. 5. 4 A graphical summary of the development of 'hybrid species distribution models' based on the community distribution models (1st correlative model) developed in this research.

H-SDMs provide a multi-step approach (Fig. 5.4) built on numerous model scenarios and simulations, mainly i) predicted scenarios of temporal changes in a given habitat (e.g. landscape dynamic models), ii) 1st correlative SDM (C-SDM), iii) development of a 2nd correlative habitat suitability model based on the environmental data changes and finally, iv) development of a mechanistic occupancy model (or population model) constrained by the predicted habitat changes (De Cáceres and Brotons, 2012). The development of such models has allowed scientists and researchers to identify possible geographic range shifts based on numerous biotic changes, such as, demography, species dispersal, functional traits and species interactions (Singer et al., 2018). While this approach is biologically and computationally intensive because of data requirements for parameterisation, the framework can use different data types at different spatial scales to predict species distribution at different grains and extents (Gavish, 2014). H-SDMs are based on the more 'traditional' SDM predictions (correlative SDMs), similar to those generated in Chapter 4 as part of this research. However, H-SDMs require an increased number of biotic variables that are known to influence the selected species (or communities). H-SDMs have weaknesses, mostly in terms of the data requirements and lack of information on paramaterisation of the models, especially data and knowledge gaps related to environmental drivers and ecological responses to such drivers. Singer et al. (2018) outlines and discusses the major challenges of developing such H-SDMs as well as providing solutions in overcoming such challenges. An additional major challenge would be to validate the results for the H-SDMs if predicted across long time-scales.

5.6 Recommendations for policy-makers: A POSTnote approach[2]

Fig. 5. 5 Photograph of 'Ashtrees Dipper' upland sheep-grazed grassland and moorland site (study site).

5.6.1 Overview

- UK upland vegetation sites contain a diverse mixture of species and vegetation types, but are threatened by climate change and agricultural practices.
- Defining and understanding the interactions between vegetation species and communities across multiple spatial scales is important to develop methods and policies for conservation.
- Understanding vegetation patch structure at small spatial scale allows scientists to predict patch changes in response to anthropogenic disturbance and can be used as indicators for future vegetation change.
- Reliable prediction of vegetation distribution provides baseline for researchers and citizen scientists on location of vegetation communities.
- Improvement of methods and models can help policy-makers to develop monitoring schemes for protecting areas of high conservation value.
- With new technologies, survey methods can be improved reducing the cost and time associated with identifying vegetation species and communities

Individual plant species do not grow in isolation, but co-occur with other plants to form vegetation communities. Too often conservation efforts have focussed on individual species, without consideration of the assemblage of other species in which they occur. Species interact with each other and the surrounding environment (also known as phytosociology) but are knowledge on some of the key environmental drivers is incomplete, especially in upland habitats. Furthermore, many of these upland habitats are threatened by anthropogenic changes. This POSTnote outlines key drivers of upland vegetation at small, medium and large spatial scales, from clumps of plants, fields through to landscapes. It also presents a generalizable approach to upland vegetation classification and methods predict their distributions in England and Wales.

5.6.2 Background

Upland habitats are typically defined as being above 250m elevation, and occupy approximately 35% of the UK land surface. Upland plant communities are affected by numerous components of the environment such as soil type, topography, meteorology, herbivore grazing, and anthropogenic disturbance (Pott, 2011). Some upland vegetation

[2] The UK's Parliamentary Office of Science and Technology (POST) regularly publishes "POSTnote" designed to inform policy-makers on key aspects of scientific research and policy. Section 5.6 is written in the style of a POSTnote, although with fewer references as it is based on the research presented in this thesis.

communities also rely on sheep and cattle grazing to sustain their development and when correctly implemented this helps to maintain a diverse mixture of different vegetation communities (Berg *et al.*, 1997; Adamson *et al.*, 2001). Unfortunately, where uplands are poorly management, for example through over-grazing, pollution, excess drainage, this may lead to declines in vegetation and animal biodiversity. Environmental change can alter relationships between neighbouring plant species and change vegetation community classes (van der Maarel, 1969). The availability of remote-sensed data (e.g. drones and satellites) can provide high resolution data to utilise and improve methods of vegetation surveys from a relatively small area to a larger landscape area. Landscape modelling of vegetation communities, using numerous variations of models, e.g. species distribution models, can quantify the relationship between the communities and the environment at different spatial scales. Exploring a bottom-up approach from vegetation patch structure through to community assemblage and to community distribution can provide several 'baselines' for policy-makers to aid development conservation measures.

5.6.3 Key environmental drivers at different spatial scales

Vegetation may appear to respond to different environmental factors depending on the scale at which it is assessed. At the individual species level, vegetation growth occurs through that of individual plants, which in upland habitats often form clumps or patches (Box 1). The shapes of these patches are influenced by, for example, soil pH, soil moisture soil composition and grazing regimes. The interplay between species and such environmental drivers can be related to plant functional traits, including root-rhizome growth, seed dispersal and regenerative properties. However, the community as a whole responds to different environmental drivers especially if assessed at larger landscape scale. Research indicates that vegetation communities, depending on their class (i.e. grassland, moorland, heathland, mires) are dependent on a multitude of large-scale variables most notably meteorological (especially temperature and rainfall) and topography. This suggests that with current climate change projections, vegetation communities will have to either adapt or undergo range shifts.

Box 1: Effect of environment on vegetation patch structure
Overall analysis of dominant and subdominant vegetation patch structure (number of patches, area and shape) indicated strong associations with the four main environmental variables of soil pH, slope, percentage water and altitude. There was weak evidence that either proximity of sheep tracks, especially within 10m of a quadrat, affected the patch structure, or that distance to the nearest drainage ditch also had an effect.

5.6.4 Application and challenges

Research into vegetation patches at a very small scale provides insights into how vegetation species interact and respond to the environment and anthropogenic disturbances. While some species can tolerate disturbance with little change in growth form, being competitors-stress tolerators (CS; see Grime et al., 1988), others undergo changes in their general patch structure, or from dominant to subdominant cover abundance. Surveying species at such a small level is an important addition to the 'toolkit' used by practising conservation scientits and research ecologists. It allows a more fine-grained assessment of the patterns in the vegetation, and may well prove to be a sensitive initial indicator of changes in the vegetation, that take longer to become apparent at larger scale, coarser-resolution surveys. By coupling such research to newly developed methods to predict vegetation community distributions at larger scales will empower researchers and policy-makers to make more informed decisions when developing and implementing environmental policies. This may include applications of different monitoring approaches such as leveraging the power of data collected through citizen science surveys more effectively. These have the potential to be both low cost and high benefit, and supplement existing databases of vegetation records.

5.6.5 Technological enhancement

Current upland vegetation research has utilised data sets at a range of spatial scales from 10cm and 1m plant and patch scale, through to field-level surveys, and finally landscape and national scales. The development of technologies for surveying purposes from for example drones through to satellite imaging, coupled with the improved processing power of modern computers and new data processing algorithms, provides a unique opportunity to increase our understanding, management and conservation of upland habitats. Drones (UAV) have the potential to take images of vegetation at fine spatial resolution, such as 1cm to 20cm scale, i.e. a high enough resolution and spatial distance to identify dominant and subdominant vegetation species, if not individual plants or leaves. Drones can only survey relatively small areas. In contrast, satellite data is available across much larger areas, often with national coverage, but at coarser spatial resolution (typically 2m to 30m) can be utilised to predict and analyse vegetation community distribution at large spatial scales (e.g. England and Wales). It is important to continue to build on the data-processing pipelines developed by (Butler, 2020) to integrate the data recorded at these different spatial scales, so that scientists have an in-depth understanding of factors affecting low-level plant growth patterns across large spatial scales.

5.6.6 Meeting the challenges

Several challenges remain to predict vegetation communities across large areas. While one method, based on modified species distribution models (SDMs), has been tested and resulted in high accuracy in model outputs, there is still a need for additional validation as well as comparison with different levels. The method used to predict communities at large spatial scale (England and Wales) has proved invaluable to identify locations in communities that are currently protected under the UK BAP. However, this method uses community-based data rather than species-level data as inputs into the SDMs. This produces spatially coarse predictions (10 km) more data intensive methods such as species stackable SDMs have the potential to predict communities at much finer spatial scales. Community and species distribution models can be time-consuming to develop, therefore an initial, relatively spatially coarse model, might be useful for policy-makers to identify areas that had not have been previously considered as conservation priorities. Newer 'hybrid' species distribution models (H-SDMs) are likely to be needed to predict range shifts in plant species and communities as a result of climate change. H-SDMs require more input data and are computationally intensive level compared to 'traditional' SDMs, but can provide predictions on both the temporal, as well as spatial, scale (Box 2). A multi-scale approach, both in time and space, to understand and assess upland vegetation species and communities is an important component to preserve and conserve natural or semi-natural environments that are under substantial anthropogenic threat. The methods and models produced as part of recent research can be both replicated, applied and improved to address research and policy gaps.

Box 2: Hybrid Species Distribution Models (H-SDMs)
H-SDMs are a multi-step modelling approach designed to integrate numerous model simulations, mainly i) predicted habitat change by climate change, ii) SDMs at numerous time steps, iii) habitat suitability model based on the environmental data changes and the output would be a population model at multiple time steps based on the predicted habitat changes driven by climate change (De Cáceres and Brotons, 2012).

5.7 Open Data Science

This research utilised numerous different data types ranging from surveyed data at 10cm scale, abundance cover survey at 1m scale, vegetation community presence data at 10km grid scale and a multitude of collections of environment data at high resolution. Data management, manipulation, analysis and prediction was done using R (R Core Team, 2013) and thus available as R scripts. Examples taken from the R scripts developed for Chapter 2 (Appendix SUP2), for

Chapter 3 (Appendix SUP3) and for Chapter 4 (Appendix SUP4) are appended as supplementary data.

In addition, to support high quality research, teaching and learning the data, results and R scripts will be made available as part of the Open Data Science scheme at Newcastle University (data.ncl). The primary aim of making the data available is to provide others with a flexible and use-friendly method access to the data resources used and algorithms developed in this research. This would also add to further work performed in this area of research by up and coming researchers, data scientists and policy makers. The availability of open data might also help researchers compare different methods of vegetation community analyses and prediction at finer scales as well as compare with different methods (especially for SDM predictions). The availability of this data would also help in international research engagement, networking and collaboration.

5.8 Final conclusions

This research aimed to understand and model upland vegetation communities at multiple spatial scales, ranging from species-level scale (10cm and 1m) through to catchment scale (Rede Catchment) and up to national scale (England and Wales). The range of data used in this research (survey data, *in situ* environmental data, aerial photographs, high resolution satellite data etc.) provided sound development of models, assessment and analysis of upland vegetation communities.

Chapter 2 in this research provided substantial information that vegetation patches at very low spatial scale (10cm and 1m) are dependent on environmental variables (both direct and indirect), and form different number of patches, areas and shapes depending on their cover type (dominant or subdominant) and growth strategy (block-formers or opportunists). Interpolation techniques can be used to develop predicted maps of patch metrics for named species at field scale. Furthermore, this research (Chapter 3) used available software and algorithms of vegetation classification, to develop robust and generalisable methods for allocating field quadrats using ordination methods. The use of ordination and hierarchical techniques makes vegetation classification generalisable to different classification systems, as well as has the potential to classify both sets of and individual quadrats where no classification is in place. Chapter 4, in this research provided insight into predicted vegetation communities using species

distribution models. The prediction of communities at both national and catchment scale provided sound information on the locations of vegetation communities that are of conservation priority, within the UK BAP. This provides both experts and citizen scientists with the knowledge on where to best survey species and communities either as part of vegetation monitoring schemes for long-term vegetation abundance data (such as the NBN Atlas or the NPMS). The predicted maps and probability of community occurrence can be used to aid identification of new potential areas for conservation management.

Chapter 6. Bibliography

Abney, S. (2002) ‘Bootstrapping’, *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics*, Philadelphia, Pennsylvania, USA, Association for Computational Linguistics, pp. 360–367 [Online]. DOI: 10.3115/1073083.1073143 (Accessed 22 May 2019).

Adamson, H. F., Ross, S. Y. and Gardner, S. M. (2001a) ‘The response of upland wet heath to ESA stocking rates: a heft scale approach’, *Biodiversity, plant structure and vegetation heterogeneity: interactions with the grazing herbivore*.

Adamson, H. F., Ross, S. Y. and Gardner, S. M. (2001b) ‘The response of upland wet heath to ESA stocking rates: a heft scale approach’, *Biodiversity, plant structure and vegetation heterogeneity: interactions with the grazing herbivore*.

Addicott, J. F., Aho, J. M., Antolin, M. F., Padilla, D. K., Richardson, J. S. and Soluk, D. A. (1987) ‘Ecological neighborhoods: scaling environmental patterns’, *Oikos*, pp. 340–346.

Aguiar, M. R. and Sala, O. E. (1999) ‘Patch structure, dynamics and implications for the functioning of arid ecosystems’, *Trends in Ecology & Evolution*, vol. 14, no. 7, pp. 273–277.

Allouche, O., Tsoar, A. and Kadmon, R. (2006) ‘Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS)’, *Journal of Applied Ecology*, vol. 43, no. 6, pp. 1223–1232 [Online]. DOI: 10.1111/j.1365-2664.2006.01214.x.

Anderson, K. and Gaston, K. J. (2013) ‘Lightweight unmanned aerial vehicles will revolutionize spatial ecology’, *Frontiers in Ecology and the Environment*, vol. 11, no. 3, pp. 138–146.

Anon (2015) ‘Fractals for Dummies’, *Bruno Marion* [Online]. Available at https://brunomarion.com/fractals-for-dummies/ (Accessed 27 November 2019).

Araújo, M. B. and Luoto, M. (2007) ‘The importance of biotic interactions for modelling species distributions under climate change’, *Global Ecology and Biogeography*, vol. 16, no. 6, pp. 743–753 [Online]. DOI: 10.1111/j.1466-8238.2007.00359.x.

Araújo, M. B. and Peterson, A. T. (2012) ‘Uses and misuses of bioclimatic envelope modeling’, *Ecology*, vol. 93, no. 7, pp. 1527–1539.

Austin, M. (2007) ‘Species distribution models and ecological theory: A critical assessment and some possible new approaches’, *Ecological Modelling*, vol. 200, no. 1–2, pp. 1–19 [Online]. DOI: 10.1016/j.ecolmodel.2006.07.005.

Austin, M. P. (2002) ‘Spatial prediction of species distribution: an interface between ecological theory and statistical modelling’, *Ecological Modelling*, vol. 157, no. 2, pp. 101–118 [Online]. DOI: 10.1016/S0304-3800(02)00205-3.

Austin, M. P. (2005) ‘Vegetation and environment: discontinuities and continuities’, *Vegetation ecology*, vol. 5284.

Austin, M. P. and Smith, T. M. (1990) ‘A new model for the continuum concept’, in *Progress in theoretical vegetation science*, Springer, pp. 35–47.

Austin, M. P. and van Niel, K. P. (2011) ‘Improving species distribution models for climate change studies: variable selection and scale’, *Journal of Biogeography*, vol. 38, no. 1, pp. 1–8.

Backshall, J., Manley, J. and Rebane, M. (2001) *The upland management handbook*, English Nature Peterborough.

Badano, E. I., Cavieres, L. A., Molina-Montenegro, M. A. and Quiroz, C. L. (2005) ‘Slope aspect influences plant association patterns in the Mediterranean matorral of central Chile’, *Journal of Arid Environments*, vol. 62, no. 1, pp. 93–108.

Ball, L. (1999) ‘Efficient rules for monetary policy’, *International finance*, vol. 2, no. 1, pp. 63–83.

Bar Massada, A., Kent, R., Blank, L., Perevolotsky, A., Hadar, L. and Carmel, Y. (2012) ‘Automated segmentation of vegetation structure units in a Mediterranean landscape’, *International Journal of Remote Sensing*, vol. 33, no. 2, pp. 346–364.

Bartlett, R. V. and Kurian, P. A. (1999) ‘The theory of environmental impact assessment: implicit models of policy making’, *Policy & Politics*, vol. 27, no. 4, pp. 415–433.

Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S. P., Peterson, A. T., Soberón, J. and Villalobos, F. (2011) ‘The crucial role of the accessible area in ecological niche modeling and species distribution modeling’, *Ecological Modelling*, vol. 222, no. 11, pp. 1810–1819.

Bates, G. H. (1935) ‘The vegetation of footpaths, sidewalks, cart-tracks and gateways’, *Journal of Ecology*, vol. 23, no. 2, pp. 470–487.

Beissinger, S. R., Iknayan, K. J., Guillera-Arroita, G., Zipkin, E. F., Dorazio, R. M., Royle, J. A. and Kéry, M. (2016) ‘Incorporating Imperfect Detection into Joint Models of Communities: A response to Warton et al.’, *Trends in Ecology & Evolution*, vol. 31, no. 10, pp. 736–737 [Online]. DOI: 10.1016/j.tree.2016.07.009.

Belgiu, M. and Drăguţ, L. (2016) ‘Random forest in remote sensing: A review of applications and future directions’, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 114, pp. 24–31.

Bennie, J., Hill, M. O., Baxter, R. and Huntley, B. (2006) 'Influence of slope and aspect on long-term vegetation change in British chalk grasslands', *Journal of ecology*, vol. 94, no. 2, pp. 355–368.

Bennie, J., Huntley, B., Wiltshire, A., Hill, M. O. and Baxter, R. (2008) 'Slope, aspect and climate: Spatially explicit and implicit models of topographic microclimate in chalk grassland', *Ecological Modelling*, vol. 216, no. 1, pp. 47–59 [Online]. DOI: 10.1016/j.ecolmodel.2008.04.010.

Berg, G., Esselink, P., Groeneweg, M. and Kiehl, K. (1997) 'Micropatterns in Festuca rubra-dominated salt-marsh vegetation induced by sheep grazing', *Plant Ecology*, vol. 132, no. 1, pp. 1–14.

Billings, W. D. (1952) 'The Environmental Complex in Relation to Plant Growth and Distribution', *The Quarterly Review of Biology*, vol. 27, no. 3, pp. 251–265.

Biodiversity Ireland (2018) *Irish Vegetation Classification* [Online]. Available at http://www.biodiversityireland.ie/projects/national-vegetation-database/irish-vegetation-classification/ (Accessed 20 February 2018).

Biron, L. (2010) 'Somerset Local Wildlife Sites And Local Geological Sites Manual Policies and Procedures for the Identification and Designation of Wildlife Sites', [Online]. Available at http://www.somerc.com/wp-content/uploads/2017/04/Somerset-Local-Sites-Guidelines-2010.pdf.

Blach-Overgaard, A., Svenning, J.-C., Dransfield, J., Greve, M. and Balslev, H. (2010) 'Determinants of palm species distributions across Africa: the relative roles of climate, non-climatic environmental factors, and spatial constraints', *Ecography*, vol. 33, no. 2, pp. 380–391 [Online]. DOI: 10.1111/j.1600-0587.2010.06273.x.

Blain, A. P. (2009) 'Understanding and Modelling Large-Scale Change in the Vegetation of the British Countryside', Newcastle University.

Blakey, R. V. and Letten, A. (2016) 'Introduction to mvabund – Environmental Computing', [Online]. Available at http://environmentalcomputing.net/introduction-to-mvabund/ (Accessed 27 November 2019).

Bonn, A., Allott, T., Hubacek, K. and Stewart, J. (2009) *Drivers of Environmental Change in Uplands*, Routledge.

Borra, S. and Di Ciaccio, A. (2010) 'Measuring the prediction error. A comparison of cross-validation, bootstrap and covariance penalty methods', *Computational statistics & data analysis*, vol. 54, no. 12, pp. 2976–2989.

Boulanger, P.-M. and Bréchet, T. (2005) 'Models for policy-making in sustainable development: The state of the art and perspectives for research', *Ecological economics*, vol. 55, no. 3, pp. 337–350.

Bradley, B. A. and Fleishman, E. (2008) 'Can remote sensing of land cover improve species distribution modelling?', *Journal of Biogeography*, vol. 35, no. 7, pp. 1158–1159.

Bradter, U., Thom, T. J., Altringham, J. D., Kunin, W. E. and Benton, T. G. (2011) 'Prediction of National Vegetation Classification communities in the British uplands using environmental data at multiple spatial scales, aerial images and the classifier random forest', *Journal of Applied Ecology*, vol. 48, no. 4, pp. 1057–1065.

Braun-Blanquet, J. (1932) 'Plant sociology. The study of plant communities. First ed.', *Plant sociology. The study of plant communities. First ed.* [Online]. Available at https://www.cabdirect.org/cabdirect/abstract/19331600801 (Accessed 3 September 2019).

Bruun, H. H., Moen, J., Virtanen, R., Grytnes, J.-A., Oksanen, L. and Angerbjörn, A. (2006) 'Effects of altitude and topography on species richness of vascular plants, bryophytes and lichens in alpine communities', *Journal of Vegetation Science*, vol. 17, no. 1, pp. 37–46 [Online]. DOI: 10.1111/j.1654-1103.2006.tb02421.x.

Bullock, J. M. and Pakeman, R. J. (1997) 'Grazing of lowland heath in England: management methods and their effects on healthland vegetation', *Biological Conservation*, vol. 79, no. 1, pp. 1–13.

Bunce, R. G. H., Barr, C. J., Gillespie, M. K., Howard, D. C., Scott, W. A., Smart, S. M., van de Poll, H. M. and Watkins, J. W. (1999) *Vegetation of the British countryside-the Countryside Vegetation System. ECOFACT Volume 1*, DETR, vol. 1.

Butler, L. (2020) 'Models of upland vegetation communities at multiple spatial scales', Newcastle-Upon-Tyne, UK, Newcastle University.

Butler, L. and Sanderson, R. A. (2018) 'Generalisable Methods for Vegetation Classification Using Computer-Generated Pseudoquadrats', *Proceedings of the iEMSs 2018 Conference*, Fort Collins, Colorado, USA, vol. 9 [Online]. Available at https://scholarsarchive.byu.edu/iemssconference/2018/.

Callcutt, K., Croft, S. and Smith, G. C. (2018) 'Predicting population trends using citizen science data: do subsampling methods produce reliable estimates for mammals?', *European Journal of Wildlife Research*, vol. 64, no. 3 [Online]. DOI: 10.1007/s10344-018-1189-7 (Accessed 4 April 2019).

Canadian National Vegetation Classification (2013) *CNVC Overview* [Online]. Available at http://cnvc-cnvc.ca/page.cfm?page=2228 (Accessed 2 February 2018).

CEH (2014) *TABLEFIT and TABLCORN* [Online]. Available at https://www.ceh.ac.uk/services/tablefit-and-tablcorn (Accessed 2 February 2018).

Chandler, M., See, L., Copas, K., Bonde, A. M. Z., López, B. C., Danielsen, F., Legind, J. K., Masinde, S., Miller-Rushing, A. J., Newman, G., Rosemartin, A. and Turak, E. (2017) 'Contribution of citizen science towards international biodiversity monitoring', *Biological Conservation*, SI:Measures of biodiversity, vol. 213, pp. 280–294 [Online]. DOI: 10.1016/j.biocon.2016.09.004.

Cherrill, A. and McClean, C. (1999) 'Between-observer variation in the application of a standard method of habitat mapping by environmental consultants in the UK', *Journal of Applied Ecology*, vol. 36, no. 6, pp. 989–1008.

Chytrý, M. and Otýpková, Z. (2003) 'Plot Sizes Used for Phytosociological Sampling of European Vegetation', *Journal of Vegetation Science*, vol. 14, no. 4, pp. 563–570.

Clements, F. E. (1907) *Plant physiology and ecology*, H. Holt.

Coulson, J. C., Butterfield, J. E. L. and Henderson, E. (1990) 'The effect of open drainage ditches on the plant and invertebrate communities of moorland and on the decomposition of peat', *Journal of Applied Ecology*, pp. 549–561.

Crain, C. M. and Bertness, M. D. (2006) 'Ecosystem Engineering across Environmental Gradients: Implications for Conservation and Management', *BioScience*, vol. 56, no. 3, pp. 211–218 [Online]. DOI: 10.1641/0006-3568(2006)056[0211:EEAEGI]2.0.CO;2.

Craine, J. M. (2005) 'Reconciling plant strategy theories of Grime and Tilman', *Journal of ecology*, vol. 93, no. 6, pp. 1041–1052.

Cranfield University (2017) 'Soils data from the National Soil Resources Institute', Cranfield University [Online]. Available at www.landis.org.uk.

Critchley, C. N. R., Chambers, B. J., Fowbert, J. A., Sanderson, R. A., Bhogal, A. and Rose, S. C. (2002) 'Association between lowland grassland plant communities and soil properties', *Biological Conservation*, vol. 105, no. 2, pp. 199–215.

Cushman, S. A., Evans, J. S., McGarigal, K. and Kiesecker, J. M. (2010) 'Toward Gleasonian Landscape Ecology: From Communities to Species, From Patches to Pixels', [Online]. DOI: 10.2737/RMRS-RP-84.

De Cáceres, M. and Brotons, L. (2012) 'Calibration of hybrid species distribution models: the value of general-purpose vs. targeted monitoring data', *Diversity and Distributions*, vol. 18, no. 10, pp. 977–989 [Online]. DOI: 10.1111/j.1472-4642.2012.00899.x.

De Marco, P., Júnior and Nóbrega, C. C. (2018) ‘Evaluating collinearity effects on species distribution models: An approach based on virtual species simulation’, *PLOS ONE*, vol. 13, no. 9, p. e0202403 [Online]. DOI: 10.1371/journal.pone.0202403.

Debinski, D. M. (2006) ‘Forest fragmentation and matrix effects: the matrix does matter’, *Journal of Biogeography*, vol. 33, no. 10, pp. 1791–1792.

DEFRA (2007) ‘Conserving Biodiversity – The UK Approach’, [Online]. Available at http://jncc.defra.gov.uk/PDF/UKBAP_ConBio-UKApproach-2007.pdf.

Dickinson, J. L., Shirk, J., Bonter, D., Bonney, R., Crain, R. L., Martin, J., Phillips, T. and Purcell, K. (2012) ‘The current state of citizen science as a tool for ecological research and public engagement’, *Frontiers in Ecology and the Environment*, vol. 10, no. 6, pp. 291–297.

Dickinson, J. L., Zuckerberg, B. and Bonter, D. N. (2010) ‘Citizen science as an ecological research tool: challenges and benefits’, *Annual review of ecology, evolution, and systematics*, vol. 41, pp. 149–172.

Dijkstra, T. M. H., Snoeren, P. R. and Gielen, C. C. A. M. (1994) ‘Extraction of three-dimensional shape from optic flow: a geometric approach’, *Journal of the Optical Society of America A*, vol. 11, no. 8, p. 2184 [Online]. DOI: 10.1364/JOSAA.11.002184.

Dodd, M. E., Silvertown, J., McConway, K., Potts, J. and Crawley, M. (1994) ‘Application of the British National Vegetation Classification to the Communities of the Park Grass Experiment through Time’, *Folia Geobotanica & Phytotaxonomica*, vol. 29, no. 3, pp. 321–334.

EDINA (2018) *Digimap* [Online]. Available at https://digimap.edina.ac.uk/environment (Accessed 9 September 2019).

Elith*, J., Graham*, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. M. M., Peterson, A. T., Phillips, S. J., Richardson, K., Scachetti-Pereira, R., Schapire, R. E., Soberón, J., Williams, S., Wisz, M. S. and Zimmermann, N. E. (2006) ‘Novel methods improve prediction of species’ distributions from occurrence data’, *Ecography*, vol. 29, no. 2, pp. 129–151 [Online]. DOI: 10.1111/j.2006.0906-7590.04596.x.

Eskildsen, A., Roux, P. C. le, Heikkinen, R. K., Høye, T. T., Kissling, W. D., Pöyry, J., Wisz, M. S. and Luoto, M. (2013) ‘Testing species distribution models across space and time: high latitude butterflies and recent warming’, *Global Ecology and Biogeography*, vol. 22, no. 12, pp. 1293–1303 [Online]. DOI: 10.1111/geb.12078.

ESRI (2015) *ArcMap | ArcGIS Desktop*, ESRI [Online]. Available at http://desktop.arcgis.com/en/arcmap/ (Accessed 15 January 2018).

European Commission (1992) *The Habitats Directive - Environment - European Commission* [Online]. Available at https://ec.europa.eu/environment/nature/legislation/habitatsdirective/index_en.htm (Accessed 21 August 2019).

European Commission (2009) *The Birds Directive - Environment - European Commission* [Online]. Available at http://ec.europa.eu/environment/nature/legislation/birdsdirective/index_en.htm (Accessed 18 June 2019).

European Commission (2010) *Strategic Environmental Assessment - SEA - Environment - European Commission* [Online]. Available at https://ec.europa.eu/environment/eia/sea-legalcontext.htm (Accessed 21 August 2019).

European Vegetation Survey (2018) *European Vegetation Archive (EVA) – European Vegetation Survey* [Online]. Available at http://euroveg.org/eva-database (Accessed 2 February 2018).

Evans, J. S., Murphy, M. A., Holden, Z. A. and Cushman, S. A. (2011) 'Modeling species distribution and change using random forest', in *Predictive species and habitat modeling in landscape ecology*, Springer, pp. 139–159.

Ewald, J. (2003) 'A critique for phytosociology', *Journal of vegetation science*, vol. 14, no. 2, pp. 291–296.

Fenton, E. W. (1937) 'The influence of sheep on the vegetation of hill grazings in Scotland', *Journal of Ecology*, vol. 25, no. 2, pp. 424–430.

Fischer, J. and Lindenmayer, D. B. (2007) 'Landscape modification and habitat fragmentation: a synthesis', *Global Ecology and Biogeography*, vol. 16, no. 3, pp. 265–280 [Online]. DOI: 10.1111/j.1466-8238.2007.00287.x.

Fisher, A. S., Podniesinski, G. S. and Leopold, D. J. (1996) 'Effects of drainage ditches on vegetation patterns in Abandoned agricultural peatlands in central New York', *Wetlands*, vol. 16, no. 4, pp. 397–409 [Online]. DOI: 10.1007/BF03161329.

Fitzpatrick, M. C., Gotelli, N. J. and Ellison, A. M. (2013) 'MaxEnt versus MaxLike: empirical comparisons with ant species distributions', *Ecosphere*, vol. 4, no. 5, pp. 1–15.

Foster, D. R., KSchoonmaker, P. and Pickett, S. T. A. (1990) 'Insights from paleoecology to community ecology', *Trends in ecology & evolution*, vol. 5, no. 4, pp. 119–122.

Fournier, A., Barbet-Massin, M., Rome, Q. and Courchamp, F. (2017) 'Predicting species distribution combining multi-scale drivers', *Global Ecology and Conservation*, vol. 12, pp. 215–226 [Online]. DOI: 10.1016/j.gecco.2017.11.002.

Franklin, J. (1995) 'Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients', *Progress in Physical Geography: Earth and Environment*, vol. 19, no. 4, pp. 474–499 [Online]. DOI: 10.1177/030913339501900403.

Franklin, J. (2010) 'Spatial point pattern analysis of plants', in *Perspectives on spatial data analysis*, Springer, pp. 113–123.

Freeman, E. A. and Moisen, G. G. (2008) 'A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa', *Ecological Modelling*, vol. 217, no. 1–2, pp. 48–58 [Online]. DOI: 10.1016/j.ecolmodel.2008.05.015.

Gao, Y., Mas, J.-F. and Navarrete, A. (2009) 'The improvement of an object-oriented classification using multi-temporal MODIS EVI satellite data', *International Journal of Digital Earth*, vol. 2, no. 3, pp. 219–236 [Online]. DOI: 10.1080/17538940902818311.

Gavish, Y. (2014) *Hybrid SDMs - Accounting for both environmental and spatial aspects in species distribution*, University of Leeds and EU BON [Online]. Available at http://biodiversity.eubon.eu/documents/20147/32722/3_EUBON_HybridSDM.pdf/839e4428-fa78-2a8f-7dbd-3a527cff3f7e.

Gleason, H. A. (1917) 'The structure and development of the plant association', *Bulletin of the Torrey Botanical Club*, vol. 44, no. 10, pp. 463–481.

Goldberg, D. E. (1985) 'Effects of Soil pH, Competition, and Seed Predation on the Distributions of Two Tree Species', *Ecology*, vol. 66, no. 2, pp. 503–511 [Online]. DOI: 10.2307/1940398.

Gower, J. C. (1975) 'Generalized procrustes analysis', *Psychometrika*, vol. 40, no. 1, pp. 33–51.

Greig-Smith, P. (1979) 'Pattern in vegetation', *Journal of Ecology*, vol. 67, no. 3, pp. 755–779.

Grime, J. P. (1988) 'The CSR model of primary plant strategies—origins, implications and tests', in *Plant evolutionary biology*, Springer, pp. 371–393.

Grime, J. P., Hodgson, J. G. and Hunt, R. (1988) 'Comparative Plant Ecology: A Functional Approach to Common British Species.(Agrostis spp., pp. 58-65.) Unwin Hyman', *London, England, UK*.

Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J. M. C., Aspinall, R. and Hastie, T. (2006) 'Making better biogeographical predictions of species' distributions', *Journal of Applied Ecology*, vol. 43, no. 3, pp. 386–392.

Guisan, A. and Thuiller, W. (2005) 'Predicting species distribution: offering more than simple habitat models', *Ecology Letters*, vol. 8, no. 9, pp. 993–1009 [Online]. DOI: 10.1111/j.1461-0248.2005.00792.x.

Guisan, A. and Zimmermann, N. E. (2000) 'Predictive habitat distribution models in ecology', *Ecological Modelling*, vol. 135, no. 2–3, pp. 147–186 [Online]. DOI: 10.1016/S0304-3800(00)00354-9.

Hampton, S. E., Strasser, C. A., Tewksbury, J. J., Gram, W. K., Budden, A. E., Batcheller, A. L., Duke, C. S. and Porter, J. H. (2013) 'Big data and the future of ecology', *Frontiers in Ecology and the Environment*, vol. 11, no. 3, pp. 156–162.

Hanberry, B. B. and He, H. S. (2013) 'Prevalence, statistical thresholds, and accuracy assessment for species distribution models', *Web Ecology*, vol. 13, no. 1, pp. 13–19 [Online]. DOI: 10.5194/we-13-13-2013.

von Hardenberg, J., Meron, E., Shachak, M. and Zarmi, Y. (2001) 'Diversity of vegetation patterns and desertification', *Physical Review Letters*, vol. 87, no. 19, p. 198101.

Hearn, S. M., Healey, J. R., McDonald, M. A., Turner, A. J., Wong, J. L. G. and Stewart, G. B. (2011) 'The repeatability of vegetation classification and mapping', *Journal of Environmental Management*, vol. 92, no. 4, pp. 1174–1184 [Online]. DOI: 10.1016/j.jenvman.2010.11.021.

Hennekens, S. M. and Schaminée, J. H. J. (2001) 'TURBOVEG, a comprehensive data base management system for vegetation data', *Journal of Vegetation Science*, vol. 12, no. 4, pp. 589–591 [Online]. DOI: 10.2307/3237010.

Hernandez, P. A., Franke, I., Herzog, S. K., Pacheco, V., Paniagua, L., Quintana, H. L., Soto, A., Swenson, J. J., Tovar, C., Valqui, T. H., Vargas, J. and Young, B. E. (2008) 'Predicting species distributions in poorly-studied landscapes', *Biodiversity and Conservation*, vol. 17, no. 6, pp. 1353–1366 [Online]. DOI: 10.1007/s10531-007-9314-z.

Hijmans, R. J., van Etten, J., Sumner, M., Cheng, J., Bevan, A., Bivand, R., Busetto, L., Canty, M., Forrest, D., Ghosh, A., Golicher, D., Gray, J., Greenberg, J. A., Hiemstra, P., Karney, C., Mattiuzzi, M., Mosher, S., Nowosad, J., Pebesma, E., Lamigueiro, O., Racine, E. B., Rowlingson, B., Shortridge, A., Venables, B. and Wueest, R. (2019) *Package 'raster'*,.

Hill, M. O. (1979a) ‘DECORANA. A Fortran program for detrended correspondence analysis and reciprocal averaging’, *Ecological and Systematics Department*.

Hill, M. O. (1979b) ‘A FORTRAN program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes’, *TWINSPAN*.

Hill, M. O. (1989) ‘Computerized Matching of Relevés and Association Tables, with an Application to the British National Vegetation Classification’, *Vegetatio*, vol. 83, no. 1/2, pp. 187–194.

Hill, M. O. (1993) ‘TABLEFIT: A program for identification of vegetation types’,.

Hill, M. O. and Gauch, H. G. (1980) ‘Detrended correspondence analysis: an improved ordination technique’, in *Classification and ordination*, Springer, pp. 47–58.

Hof, A. R., Jansson, R. and Nilsson, C. (2012) ‘The usefulness of elevation as a predictor variable in species distribution modelling’, *Ecological Modelling*, vol. 246, pp. 86–90 [Online]. DOI: 10.1016/j.ecolmodel.2012.07.028.

Holden, J., Chapman, P. J. and Labadz, J. C. (2004) ‘Artificial drainage of peatlands: hydrological and hydrochemical process and wetland restoration’, *Progress in Physical Geography*, vol. 28, no. 1, pp. 95–123 [Online]. DOI: 10.1191/0309133304pp403ra.

Hui, F. K., Taskinen, S., Pledger, S., Foster, S. D. and Warton, D. I. (2015) ‘Model-based approaches to unconstrained ordination’, *Methods in Ecology and Evolution*, vol. 6, no. 4, pp. 399–411.

Jackson, S. F. and Gaston, K. J. (2008) ‘Land use change and the dependence of national priority species on protected areas’, *Global Change Biology*, vol. 14, no. 9, pp. 2132–2138 [Online]. DOI: 10.1111/j.1365-2486.2008.01628.x.

Jiménez-Valverde, A. (2012) ‘Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling’, *Global Ecology and Biogeography*, vol. 21, no. 4, pp. 498–507 [Online]. DOI: 10.1111/j.1466-8238.2011.00683.x.

JNCC (2009) *NVC survey data & distribution maps* [Online]. Available at http://archive.jncc.gov.uk/default.aspx?page=4267 (Accessed 10 September 2019).

JNCC (2016) *UK BAP list of habitats* [Online]. Available at http://jncc.defra.gov.uk/page-5706 (Accessed 16 January 2018).

JNCC (2019) *UK BAP | JNCC - Adviser to Government on Nature Conservation* [Online]. Available at https://jncc.gov.uk/our-work/uk-bap/ (Accessed 10 September 2019).

JNCC and DEFRA (2012) *UK Post-2010 Biodiversity Framework*,.

Jongman, E. (1995) *Data analysis in community and landscape ecology*, Cambridge university press.

Jongman, R. H. G., Ter Braak, C. J. F. and van Tongeren, O. F. (1987) *R. 1995. Data analysis in community and landscape ecology*, Cambridge Univ. Press.

Kampichler, C., Wieland, R., Calmé, S., Weissenberger, H. and Arriaga-Weiss, S. (2010) 'Classification in conservation biology: a comparison of five machine-learning methods', *Ecological Informatics*, vol. 5, no. 6, pp. 441–450.

Ke, Y., Im, J., Lee, J., Gong, H. and Ryu, Y. (2015) 'Characteristics of Landsat 8 OLI-derived NDVI by comparison with multiple satellite sensors and in-situ observations', *Remote Sensing of Environment*, vol. 164, pp. 298–313 [Online]. DOI: 10.1016/j.rse.2015.04.004.

Kent, M., Gill, W. J., Weaver, R. E. and Armitage, R. P. (1997) 'Landscape and plant community boundaries in biogeography', *Progress in physical geography*, vol. 21, no. 3, pp. 315–353.

Kikvidze, Z. and Ohsawa, M. (2002) 'Measuring the number of co-dominants in ecological communities', *Ecological Research*, vol. 17, no. 4, pp. 519–525 [Online]. DOI: 10.1046/j.1440-1703.2002.00508.x.

Klausmeier, C. A. (1999) 'Regular and irregular patterns in semiarid vegetation', *Science*, vol. 284, no. 5421, pp. 1826–1828.

Krahulec, F., Skálová, H., Herben, T., Hadincová, V., Wildová, R. and Pecháčková, S. (2001) 'Vegetation changes following sheep grazing in abandoned mountain meadows', *Applied Vegetation Science*, vol. 4, no. 1, pp. 97–102 [Online]. DOI: 10.1111/j.1654-109X.2001.tb00239.x.

Kremen, C. (2005) 'Managing ecosystem services: what do we need to know about their ecology?', *Ecology Letters*, vol. 8, no. 5, pp. 468–479 [Online]. DOI: 10.1111/j.1461-0248.2005.00751.x.

Lambin, X., Cornulier, T., Oliver, M. K. and Fraser, E. J. (2014) *Analysis and future application of Hebridean Mink Project data*, SNH.

Landcare Research (2016) - *National Vegetation Survey* [Online]. Available at https://nvs.landcareresearch.co.nz/ (Accessed 2 February 2018).

Lange, R. T. (1969) 'The piosphere: sheep track and dung patterns', *Journal of Range Management*, pp. 396–400.

Lausch, A., Pause, M., Merbach, I., Zacharias, S., Doktor, D., Volk, M. and Seppelt, R. (2013) 'A new multiscale approach for monitoring vegetation using remote sensing-

based indicators in laboratory, field, and landscape', *Environmental monitoring and assessment*, vol. 185, no. 2, pp. 1215–1235.

Lawrence, A. (2006) '"No Personal Motive?" Volunteers, Biodiversity, and the False Dichotomies of Participation', *Ethics, Place & Environment*, vol. 9, no. 3, pp. 279–298 [Online]. DOI: 10.1080/13668790600893319.

Lawrence, A. (2010) *Taking Stock of Nature: Participatory Biodiversity Assessment for Policy, Planning and Practice*, Cambridge University Press.

Lawson, C. R., Hodgson, J. A., Wilson, R. J. and Richards, S. A. (2013) 'Prevalence, thresholds and the performance of presence-absence models', Freckleton, R. (ed), *Methods in Ecology and Evolution*, vol. 5, no. 1, pp. 54–64 [Online]. DOI: 10.1111/2041-210X.12123.

Legendre, P. and Gallagher, E. D. (2001) 'Ecologically meaningful transformations for ordination of species data', *Oecologia*, vol. 129, no. 2, pp. 271–280 [Online]. DOI: 10.1007/s004420100716.

Legendre, P. and Legendre, L. F. (2012) *Numerical ecology*, Elsevier.

Lepš, J. and Šmilauer, P. (2003) *Multivariate analysis of ecological data using CANOCO*, Cambridge university press.

Li, B.-L. (2000) 'Fractal geometry applications in description and analysis of patch patterns and patch dynamics', *Ecological Modelling*, vol. 132, no. 1–2, pp. 33–50.

Li, W., Goodchild, M. F. and Church, R. (2013) 'An efficient measure of compactness for two-dimensional shapes and its application in regionalization problems', *International Journal of Geographical Information Science*, vol. 27, no. 6, pp. 1227–1250 [Online]. DOI: 10.1080/13658816.2012.752093.

Liddle, M. J. and Greig-Smith, P. (1975) 'A survey of tracks and paths in a sand dune ecosystem', *Journal of Applied Ecology*, vol. 12, no. 3, pp. 893–908.

Liu, C., White, M. and Newell, G. (2009) 'Measuring the accuracy of species distribution models: a review', *Proceedings 18th World IMACs/MODSIM Congress. Cairns, Australia*, pp. 4241–4247.

Liu, C., White, M. and Newell, G. (2011) 'Measuring and comparing the accuracy of species distribution models with presence-absence data', *Ecography*, vol. 34, no. 2, pp. 232–243 [Online]. DOI: 10.1111/j.1600-0587.2010.06354.x.

Liu, C.-R. and CHEN, L.-Z. (2000) 'Landscape scale fractal analysis of patch shape in the vegetation of the Beijing region', *Acta Phytoecologica Sinica*, vol. 24, no. 2, pp. 129–134.

Lomolino, M. V. (2001) ‘Elevation gradients of species-density: historical and prospective views’, *Global Ecology and Biogeography*, vol. 10, no. 1, pp. 3–13 [Online]. DOI: 10.1046/j.1466-822x.2001.00229.x.

Malik, R. N. and Husain, S. Z. (2006) ‘Classification and ordination of vegetation communities of the Lohibehr reserve forest and its surrounding areas, Rawalpindi, Pakistan’, *Pakistan Journal of Botany*, vol. 38, no. 3, p. 543.

Malloch, A. J. C. (1998) ‘A computer program to aid the assignment of vegetation data to the communities and subcommunities of the National Vegetation Classification’, *University of Lancaster, UK*.

Manel, S., Williams, H. C. and Ormerod, S. J. (2001) ‘Evaluating presence–absence models in ecology: the need to account for prevalence’, *Journal of Applied Ecology*, vol. 38, no. 5, pp. 921–931 [Online]. DOI: 10.1046/j.1365-2664.2001.00647.x.

Måren, I. E., Karki, S., Prajapati, C., Yadav, R. K. and Shrestha, B. B. (2015) ‘Facing north or south: Does slope aspect impact forest stand characteristics and soil properties in a semiarid trans-Himalayan valley?’, *Journal of Arid Environments*, vol. 121, pp. 112–123 [Online]. DOI: 10.1016/j.jaridenv.2015.06.004.

Maron, J. L. and Crone, E. (2006) ‘Herbivory: effects on plant abundance, distribution and population growth’, *Proceedings of the Royal Society B: Biological Sciences*, vol. 273, no. 1601, pp. 2575–2584 [Online]. DOI: 10.1098/rspb.2006.3587.

Masek, J. (2019) *Landsat 8 « Landsat Science* [Online]. Available at https://landsat.gsfc.nasa.gov/landsat-data-continuity-mission/ (Accessed 21 June 2019).

McGarigal, K. (2017) ‘Landscape metrics for categorical map patterns’, *Lecture Notes. Available online: http://www. umass. edu/landeco/teaching/landscape_ecology/schedule/chapter9_metrics. pdf (accessed on 3 July 2018)*.

McLaughlin, A. and Mineau, P. (1995) ‘The impact of agricultural practices on biodiversity’, *Agriculture, Ecosystems & Environment*, vol. 55, no. 3, pp. 201–212 [Online]. DOI: 10.1016/0167-8809(95)00609-V.

McLeod, C., Yeo, M., Brown, A., Burn, A., Hopkins, J. and Way, S. (2009) ‘The Habitats Directive: selection of Special Areas of Conservation in the UK’, p. 34.

Milchunas, D. G. and Lauenroth, W. K. (1993) ‘Quantitative effects of grazing on vegetation and soils over a global range of environments: Ecological Archives M063-001’, *Ecological monographs*, vol. 63, no. 4, pp. 327–366.

Miles, J. (1981) ‘Problems in heathland and grassland dynamics’, *Vegetatio*, vol. 46–47, no. 1, pp. 61–74 [Online]. DOI: 10.1007/BF00118384.

Miller, J. R. and Hobbs, R. J. (2007) ‘Habitat Restoration?Do We Know What We?re Doing?’, *Restoration Ecology*, vol. 15, no. 3, pp. 382–390 [Online]. DOI: 10.1111/j.1526-100X.2007.00234.x.

Milne, J. A., Pakeman, R. J., Kirkham, F. W., Jones, I. P. and Hossell, J. E. (2002) ‘Biomass production of upland vegetation types in England and Wales’, *Grass and Forage Science*, vol. 57, no. 4, pp. 373–388.

Morecroft, M. D., Bealey, C. E., Beaumont, D. A., Benham, S., Brooks, D. R., Burt, T. P., Critchley, C. N. R., Dick, J., Littlewood, N. A. and Monteith, D. T. (2009) ‘The UK Environmental Change Network: emerging trends in the composition of plant and animal communities and the physical environment’, *Biological Conservation*, vol. 142, no. 12, pp. 2814–2832.

Mountrakis, G., Im, J. and Ogole, C. (2011) ‘Support vector machines in remote sensing: A review’, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 3, pp. 247–259 [Online]. DOI: 10.1016/j.isprsjprs.2010.11.001.

Naimi, B. and Araújo, M. B. (2016) ‘sdm: a reproducible and extensible R platform for species distribution modelling’, *Ecography*, vol. 39, no. 4, pp. 368–375 [Online]. DOI: 10.1111/ecog.01881.

Narkhede, S. (2018) *Understanding Confusion Matrix* [Online]. Available at https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62 (Accessed 27 November 2019).

NASA (2019) ‘Thermal Infrared Sensor (TIRS) « Landsat Science’, [Online]. Available at https://landsat.gsfc.nasa.gov/thermal-infrared-sensor-tirs/ (Accessed 10 September 2019).

National Biodiversity Network (2019) *National Biodiversity Network* [Online]. Available at https://nbn.org.uk/ (Accessed 21 June 2019).

Ohsawa, M. (1984) ‘Differentiation of vegetation zones and species strategies in the subalpine region of Mt. Fuji’, *Vegetatio*, vol. 57, no. 1, pp. 15–52 [Online]. DOI: 10.1007/BF00031929.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O’hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H. and Wagner, H. (2015) ‘Package “vegan”’, *Community ecology package, version*, vol. 2, no. 9.

Pan, W. (1999) ‘Bootstrapping likelihood for model selection with small samples’, *Journal of Computational and Graphical Statistics*, vol. 8, no. 4, pp. 687–698.

Parkes, D., Newell, G. and Cheal, D. (2003) 'Assessing the quality of native vegetation: The "habitat hectares" approach', *Ecological Management & Restoration*, vol. 4, no. s1, pp. S29–S38 [Online]. DOI: 10.1046/j.1442-8903.4.s.4.x.

Pärtel, M., Helm, A., Ingerpuu, N., Reier, Ü. and Tuvi, E.-L. (2004) 'Conservation of Northern European plant diversity: the correspondence with soil pH', *Biological Conservation*, vol. 120, no. 4, pp. 525–531 [Online]. DOI: 10.1016/j.biocon.2004.03.025.

Pateman, R. and Hodgson, J. (2012) 'Range Margins and Climates', *NERC*, p. 43.

Pebesma, E., Bivand, R., Racine, E., Sumner, M., Cook, I., Keitt, T., Lovelace, R., Wickham, H., Ooms, J., Muller, K. and Pederson, T. L. (2019) *Package 'sf'*,.

Pescott, O. L., Walker, K. J., Harris, F., New, H., Cheffings, C. M., Newton, N., Jitlal, M., Redhead, J., Smart, S. M. and Roy, D. B. (2019) 'The design, launch and assessment of a new volunteer-based plant monitoring scheme for the United Kingdom', *PloS one*, vol. 14, no. 4, p. e0215891.

Pescott, O. L., Walker, K. J., Pocock, M. J., Jitlal, M., Outhwaite, C. L., Cheffings, C. M., Harris, F. and Roy, D. B. (2015) 'Ecological monitoring with citizen science: the design and implementation of schemes for recording plants in Britain and Ireland', *Biological Journal of the Linnean Society*, vol. 115, no. 3, pp. 505–521.

Phillips, S. J., Dudík, M. and Schapire, R. E. (2004) 'A maximum entropy approach to species distribution modeling', *Proceedings of the twenty-first international conference on Machine learning*, ACM, p. 83.

Plumptre, A. J. (1994) 'The effects of trampling damage by herbivores on the vegetation of the Pare National des Volcans, Rwanda', *African Journal of Ecology*, vol. 32, no. 2, pp. 115–129.

Pollet, M., Meuffels, H. and Grootaert, P. (2001) 'Status of Dolichopodidae of the Flemish red data book in the Netherlands (Insecta: Diptera)', *Changes in ranges*, vol. 2, p. 61.

Pott, R. (2011) 'Phytosociology: A modern geobotanical method', *Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology*, vol. 145, no. sup1, pp. 9–18.

Pouteau, R., Meyer, J.-Y., Taputuarai, R. and Stoll, B. (2012) 'Support vector machines to map rare and endangered native plants in Pacific islands forests', *Ecological Informatics*, vol. 9, pp. 37–46 [Online]. DOI: 10.1016/j.ecoinf.2012.03.003.

Pywell, R. F., Meek, W. R., Loxton, R. G., Nowakowski, M., Carvell, C. and Woodcock, B. A. (2011) 'Ecological restoration on farmland can drive beneficial functional

responses in plant and invertebrate communities', *Agriculture, Ecosystems & Environment*, vol. 140, no. 1, pp. 62–67 [Online]. DOI: 10.1016/j.agee.2010.11.012.

Ramchunder, S. J., Brown, L. E. and Holden, J. (2009) 'Environmental effects of drainage, drain-blocking and prescribed vegetation burning in UK upland peatlands', *Progress in Physical Geography*, vol. 33, no. 1, pp. 49–79 [Online]. DOI: 10.1177/0309133309105245.

Raymond, B., van der Wal, J. and Belbin, L. (2019) *NBN4R: National Biodiversity Network Atlas (UK and Ireland) Data and Resource in R*, [Online]. Available at https://rdrr.io/github/fozy81/NBN4R/.

Ritchie, M. E. (2009) *Scale, heterogeneity, and the structure and diversity of ecological communities*, Princeton University Press, vol. 47.

Robinson, D. and Rorison, I. H. (1983) 'A comparison of the responses of Lolium perenne L.*, Holcus lanatus L. and Deschampsia flexuosa (L.) Trin. To a localised supply of nitrogen', *New Phytologist*, vol. 94, no. 2, pp. 263–273.

Rodwell, J. S. (1998a) *British Plant Communities Volume 1. Grassland and montane communities*, Cambridge University Press.

Rodwell, J. S. (1998b) *British plant communities. Volume 3. Grassland and montane communities*,.

Rodwell, John S. (1998) *British plant communities Volume 2. Grassland and montane communities*, Cambridge University Press, vol. 2.

Rodwell, J. S. (2006) *National vegetation classification: users' handbook*, Joint nature conservation committee.

Roleček, J., Tichý, L., Zelený, D. and Chytrý, M. (2009) 'Modified TWINSPAN classification in which the hierarchy respects cluster heterogeneity', *Journal of Vegetation Science*, vol. 20, no. 4, pp. 596–602.

Ross, L. C., Woodin, S. J., Hester, A. J., Thompson, D. B. A. and Birks, H. J. B. (2012) 'Biotic homogenization of upland vegetation: patterns and drivers at multiple spatial scales over five decades', *Journal of Vegetation Science*, vol. 23, no. 4, pp. 755–770 [Online]. DOI: 10.1111/j.1654-1103.2012.01390.x.

le Roux, P. C., Pellissier, L., Wisz, M. S. and Luoto, M. (2014) 'Incorporating dominant species as proxies for biotic interactions strengthens plant community models', *Journal of Ecology*, vol. 102, no. 3, pp. 767–775.

Rowland, C. S., Morton, R. D., Carrasco, L., McShane, G., O'Neil, A. W. and Wood, C. M. (2017) 'Land Cover Map 2015 (vector, GB)', NERC Environmental Information Data

Centre [Online]. Available at https://doi.org/10.5285/6c6c9203-7333-4d96-88ab-78925e7a4e73.

Royle, J. A., Chandler, R. B., Yackulic, C. and Nichols, J. D. (2012) 'Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions', *Methods in Ecology and Evolution*, vol. 3, no. 3, pp. 545–554 [Online]. DOI: 10.1111/j.2041-210X.2011.00182.x.

RSPB (2015) *The Uplands: Time to Change?* [Online]. Available at http://ww2.rspb.org.uk/Images/uplands_tcm9-166286.pdf (Accessed 28 May 2019).

Rushton, S. P. (1992) 'A preliminary model for investigating the ecological consequences of land use change within the framework of the ITE land classification', *Land use change: the causes and consequences*, no. 27, p. 111.

Rushton, S. P., Byrne, J. P. and Young, A. G. (1992) 'Redesdale: a preliminary test catchment for the NERC/ESRC Land Use Programme', *ITE symposium*.

Rushton, S. P., Ormerod, S. J. and Kerby, G. (2004) 'New paradigms for modelling species distributions?', *Journal of applied ecology*, vol. 41, no. 2, pp. 193–200.

RYAN, R. (2018) 'FROM THE NBN ATLAS', *HEMIPTERIST*, p. 191.

Sales, A. T., Gomes, M. G., de Sousa, W. H., Cezar, M. F. and Souto, D. V. O. (2016) 'Performance of goats and sheep grazing in Brazilian semi-arid scrubland supplemented with feed-blocks', vol. 24, p. 7.

Sanderson, R. A. and Rushton, S. P. (1995) 'VEMM: predicting the effects of agricultural management and environmental conditions on semi-natural vegetation', *Computers and electronics in agriculture*, vol. 12, no. 3, pp. 237–247.

Sanderson, R. A., Rushton, S. P., Cherrill, A. J. and Byrne, J. P. (1995) 'Soil, vegetation and space: an analysis of their effects on the invertebrate communities of a moorland in north-east England', *Journal of Applied Ecology*, pp. 506–518.

Sanderson, R. A., Rushton, S. P., Pickering, A. T. and Byrne, J. P. (1995) 'A preliminary method of predicting plant species distributions using the British National Vegetation Classification', *Journal of environmental management*, vol. 43, no. 3, pp. 265–288.

Sarparast, M. (2018) *LSRS: Land Surface Remote Sensing*, [Online]. Available at https://cran.r-project.org/web/packages/LSRS/index.html.

Seton-Clements, E., Stewart, G., Close, A. and Sanderson, R. (in prep) 'Is scientific policy guidance reliable?',.

Silvertown, J. (2009) 'A new dawn for citizen science', *Trends in Ecology & Evolution*, vol. 24, no. 9, pp. 467–471 [Online]. DOI: 10.1016/j.tree.2009.03.017.

Sims, J. T. (1986) ‘Soil pH Effects on the Distribution and Plant Availability of Manganese, Copper, and Zinc 1’, *Soil Science Society of America Journal*, vol. 50, no. 2, pp. 367–373 [Online]. DOI: 10.2136/sssaj1986.03615995005000020023x.

Sinclair, S., White, M. and Newell, G. (2010) ‘How useful are species distribution models for managing biodiversity under future climates?’, *Ecology and Society*, vol. 15, no. 1.

Singer, A., Schweiger, O., Kühn, I. and Johst, K. (2018) ‘Constructing a hybrid species distribution model from standard large-scale distribution data’, *Ecological Modelling*, vol. 373, pp. 39–52 [Online]. DOI: 10.1016/j.ecolmodel.2018.02.002.

Smart, S., Goodwin, A., Wallace, H. and Jones, M. (2016) ‘Modular Analysis of Vegetation Information System (MAVIS)’, [Online]. Available at https://www.ceh.ac.uk/services/modular-analysis-vegetation-information-system-mavis (Accessed 15 January 2018).

Smith, A. B. and Santos, M. J. (2019) ‘Testing the ability of species distribution models to infer variable importance’, *bioRxiv* [Online]. DOI: 10.1101/715904 (Accessed 26 November 2019).

Smith, G. M. and Wyatt, B. K. (2007a) ‘Multi-scale survey by sample-based field methods and remote sensing: A comparison of UK experience with European environmental assessments’, *Landscape and Urban Planning*, vol. 79, no. 2, pp. 170–176 [Online]. DOI: 10.1016/j.landurbplan.2006.02.011.

Smith, G. M. and Wyatt, B. K. (2007b) ‘Multi-scale survey by sample-based field methods and remote sensing: a comparison of UK experience with European environmental assessments’, *Landscape and Urban Planning*, vol. 79, no. 2, pp. 170–176.

Smith, P. A. (1996) ‘Botanical recording, distribution maps and species frequency’,.

Smith, R. S., Charman, D., Rushton, S. P., Sanderson, R. A., Simkin, J. M. and Shiel, R. S. (2003) ‘Vegetation change in an ombrotrophic mire in northern England after excluding sheep’, *Applied Vegetation Science*, vol. 6, no. 2, pp. 261–270.

Smith, R. S., Rushton, S. P. and Wadesworth, R. A. (1992) ‘Predicting vegetation change in an upland environment’, *ITE symposium*.

Stage, A. R. and Salas, C. (2007) ‘Interactions of elevation, aspect, and slope in models of forest species composition and productivity’, *Forest Science*, vol. 53, no. 4, pp. 486–492.

Stewart, G. B. and Pullin, A. S. (2006) ‘Does sheep-grazing degrade unimproved neutral grasslands managed as pasture in lowland Britain’, *Systematic Review*, vol. 15.

Stewart, G. B. and Pullin, A. S. (2008) ‘The relative importance of grazing stock type and grazing intensity for conservation of mesotrophic “old meadow” pasture’, *Journal for*

Nature Conservation, vol. 16, no. 3, pp. 175–185 [Online]. DOI: 10.1016/j.jnc.2008.09.005.

Stewart, G. B. and Schmid, C. H. (2015) 'Lessons from meta-analysis in ecology and evolution: the need for trans-disciplinary evidence synthesis methodologies', *Research Synthesis Methods*, vol. 6, no. 2, pp. 109–110 [Online]. DOI: 10.1002/jrsm.1152.

STFC (2018) *Dataset Collection Record* [Online]. Available at https://catalogue.ceda.ac.uk/uuid/4dc8450d889a491ebb20e724debe2dfb (Accessed 10 September 2019).

Suding, K. N., Lavorel, S., Chapin Iii, F. S., Cornelissen, J. H., DIAz, S., Garnier, E., Goldberg, D., Hooper, D. U., Jackson, S. T. and NAVAS, M.-L. (2008) 'Scaling environmental change through the community-level: a trait-based response-and-effect framework for plants', *Global Change Biology*, vol. 14, no. 5, pp. 1125–1140.

Sugihara, G. and May, R. M. (1990) 'Applications of fractals in ecology', *Trends in Ecology & Evolution*, vol. 5, no. 3, pp. 79–86.

Team, R. C. (2013) 'R: A language and environment for statistical computing',.

Ter Braak, C. J. F. (1987) 'The analysis of vegetation-environment relationships by canonical correspondence analysis', *Vegetatio*, vol. 69, no. 1, pp. 69–77 [Online]. DOI: 10.1007/BF00038688.

Thomas, A. S. (1959) 'SHEEP PATHSO: Observations on the variability of chalk pastures', *Grass and Forage Science*, vol. 14, no. 3, pp. 157–164.

Thompson, D. B. A., MacDonald, A. J., Marsden, J. H. and Galbraith, C. A. (1995) 'Upland heather moorland in Great Britain: A review of international importance, vegetation change and some objectives for nature conservation', *Biological Conservation*, vol. 71, no. 2, pp. 163–178 [Online]. DOI: 10.1016/0006-3207(94)00043-P.

Thuiller, W., Araujo, M. B. and Lavorel, S. (2004) 'Do we need land-cover data to model species distributions in Europe?', *Journal of Biogeography*, vol. 31, no. 3, pp. 353–361.

Tscharntke, T., Klein, A. M., Kruess, A., Steffan-Dewenter, I. and Thies, C. (2005) 'Landscape perspectives on agricultural intensification and biodiversity – ecosystem service management', *Ecology Letters*, vol. 8, no. 8, pp. 857–874 [Online]. DOI: 10.1111/j.1461-0248.2005.00782.x.

Tulloch, A. I., Possingham, H. P., Joseph, L. N., Szabo, J. and Martin, T. G. (2013) 'Realising the full potential of citizen science monitoring programs', *Biological Conservation*, vol. 165, pp. 128–138.

United Nations (1992) ‘Convention on Biological Diversity’, [Online]. Available at https://www.cbd.int/doc/legal/cbd-en.pdf.

U.S. Geological Survey, Earth Resources Observation And Science Center (2012) ‘Collection-1 Landsat ETM+ Level-2 Surface Reflectance (SR) Science Product’, U.S. Geological Survey [Online]. DOI: 10.5066/F7Q52MNK (Accessed 3 April 2019).

USGS (2010) *GMTED2010* [Online]. Available at https://topotools.cr.usgs.gov/GMTED_viewer/gmted2010_fgdc_metadata.html (Accessed 21 June 2019).

USNVC (2016) ‘Explore The Classification « The U.S. National Vegetation Classification’, [Online]. Available at http://usnvc.org/explore-classification/ (Accessed 2 February 2018).

van der Maarel, E. (1969) ‘On the use of ordination models in phytosociology’, *Vegetatio*, vol. 19, no. 1–6, pp. 21–46.

van der Maarel, E. (1979) ‘Transformation of cover-abundance values in phytosociology and its effects on community similarity’, *Vegetatio*, vol. 39, no. 2, pp. 97–114.

van der Maarel, E. (1988) ‘Vegetation dynamics: patterns in time and space’, in *Temporal and Spatial Patterns of Vegetation Dynamics*, Springer, pp. 7–19.

van der Maarel, E. (1996) ‘Pattern and process in the plant community: Fifty years after AS Watt’, *Journal of Vegetation Science*, vol. 7, no. 1, pp. 19–28.

van Strien, A. J., van der Burg, T., Rip, W. J. and Strucker, R. C. W. (1991) ‘Effects of Mechanical Ditch Management on the Vegetation of Ditch Banks in Dutch Peat Areas’, *Journal of Applied Ecology*, vol. 28, no. 2, pp. 501–513 [Online]. DOI: 10.2307/2404564.

Vesk, P. A. and Westoby, M. (2001) ‘Predicting Plant Species’ Responses to Grazing’, *Journal of Applied Ecology*, vol. 38, no. 5, pp. 897–909.

Wang, X., Gao, Q., Wang, C. and Yu, M. (2017) ‘Spatiotemporal patterns of vegetation phenology change and relationships with climate in the two transects of East China’, *Global Ecology and Conservation*, vol. 10, pp. 206–219 [Online]. DOI: 10.1016/j.gecco.2017.01.010.

Wang, Y., Naumann, U., Eddelbuettel, D., Wilshire, J. and Warton, D. (2019) *mvabund: Statistical methods for analysing multivariate abundance data*, [Online]. Available at https://cran.r-project.org/web/packages/mvabund/index.html.

Warton, D. I., Blanchet, F. G., O’Hara, R. B., Ovaskainen, O., Taskinen, S., Walker, S. C. and Hui, F. K. (2015) ‘So many variables: joint modeling in community ecology’, *Trends in Ecology & Evolution*, vol. 30, no. 12, pp. 766–779.

Warton, D. I., Blanchet, F. G., O'Hara, R., Ovaskainen, O., Taskinen, S., Walker, S. C. and Hui, F. K. C. (2016) 'Extending Joint Models in Community Ecology: A Response to Beissinger et al.', *Trends in Ecology & Evolution*, vol. 31, no. 10, pp. 737–738 [Online]. DOI: 10.1016/j.tree.2016.07.007.

Warton, D. I., Wright, S. T. and Wang, Y. (2012) 'Distance-based multivariate analyses confound location and dispersion effects', *Methods in Ecology and Evolution*, vol. 3, no. 1, pp. 89–101 [Online]. DOI: 10.1111/j.2041-210X.2011.00127.x.

Watt, A. S. (1947) 'Pattern and process in the plant community', *Journal of ecology*, vol. 35, no. 1/2, pp. 1–22.

White, P. S. (1979) 'Pattern, process, and natural disturbance in vegetation', *The botanical review*, vol. 45, no. 3, pp. 229–299.

Whittaker, R. H. (1965) 'Dominance and diversity in land plant communities: numerical relations of species express the importance of competition in community function and evolution', *Science*, vol. 147, no. 3655, pp. 250–260.

Wickham, H. (2016) *ggplot2: Elegant graphics for data analysis.*, New York, Springer-Verlag.

Wigmosta, M. S., Nijssen, B., Storck, P. and Lettenmaier, D. P. (2002) 'The distributed hydrology soil vegetation model', *Mathematical models of small watershed hydrology and applications*, pp. 7–42.

Xiaoqin, W., Miaomiao, W., Shaoqiang, W. and Yundong, W. (2015) 'Extraction of vegetation information from visible unmanned aerial vehicle images.', *Transactions of the Chinese Society of Agricultural Engineering*, vol. 31, no. 5.

Zak, M. R. and Cabido, M. (2002) 'Spatial patterns of the Chaco vegetation of central Argentina: integration of remote sensing and phytosociology', *Applied Vegetation Science*, vol. 5, no. 2, pp. 213–226.

Zelený, D. (2017) *Analysis of community ecology data in R*,.

Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C. F., Gao, F., Reed, B. C. and Huete, A. (2003) 'Monitoring vegetation phenology using MODIS', *Remote Sensing of Environment*, vol. 84, no. 3, pp. 471–475 [Online]. DOI: 10.1016/S0034-4257(02)00135-9.

Zhang, X., Wang, J., Gao, F., Liu, Y., Schaaf, C., Friedl, M., Yu, Y., Jayavelu, S., Gray, J., Liu, L., Yan, D. and Henebry, G. M. (2017) 'Exploration of scaling effects on coarse resolution land surface phenology', *Remote Sensing of Environment*, vol. 190, pp. 318–330 [Online]. DOI: 10.1016/j.rse.2017.01.001.

Appendix 1. Patch metrics

Appendix 1.1. Maps of Environmental Variables at Ashtrees

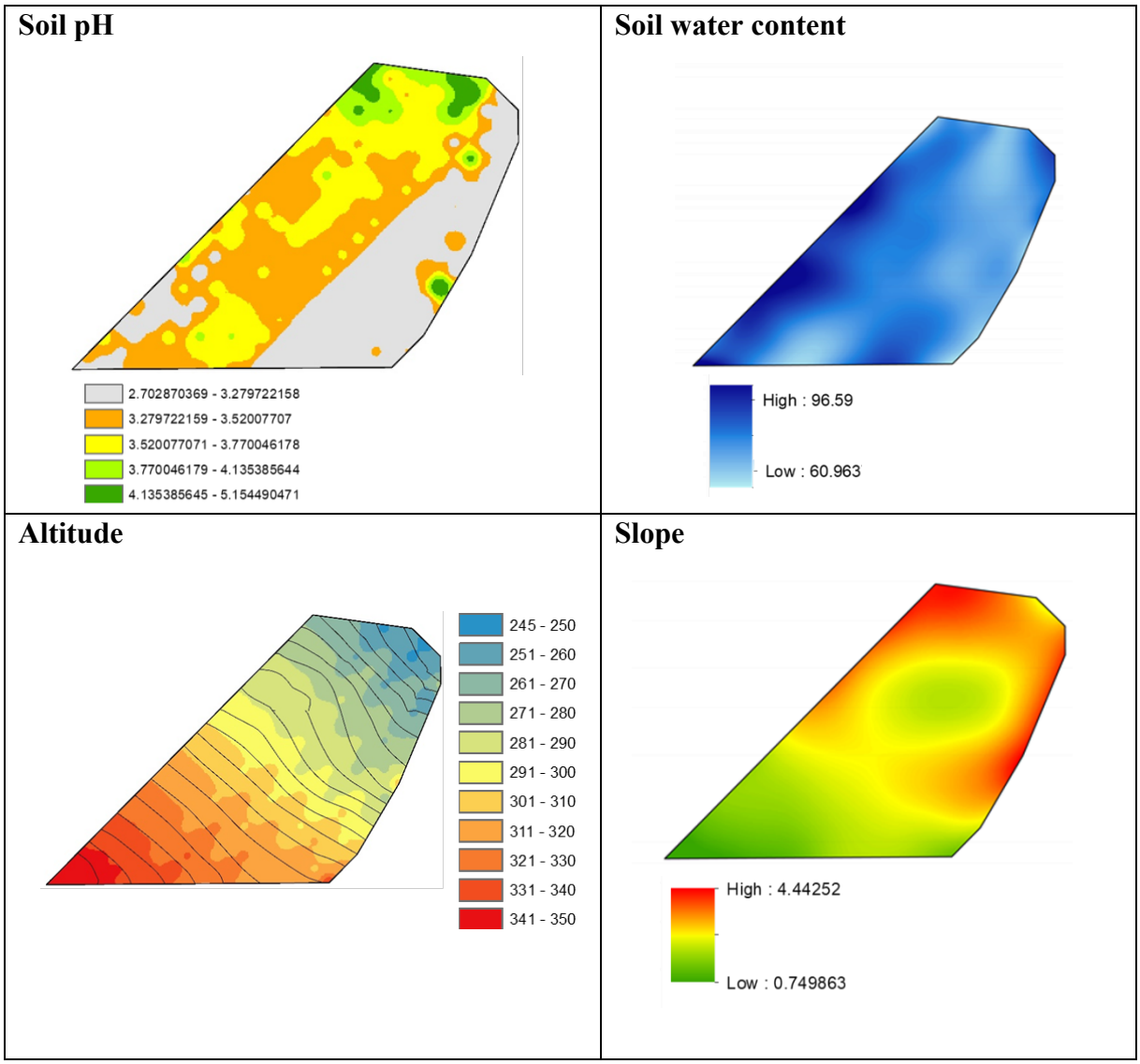


Fig. S2. 1 Interpolated maps of soil pH, soil water content, altitude and slope across Ashtrees Dipper.

Appendix 1.2. Number of Patches

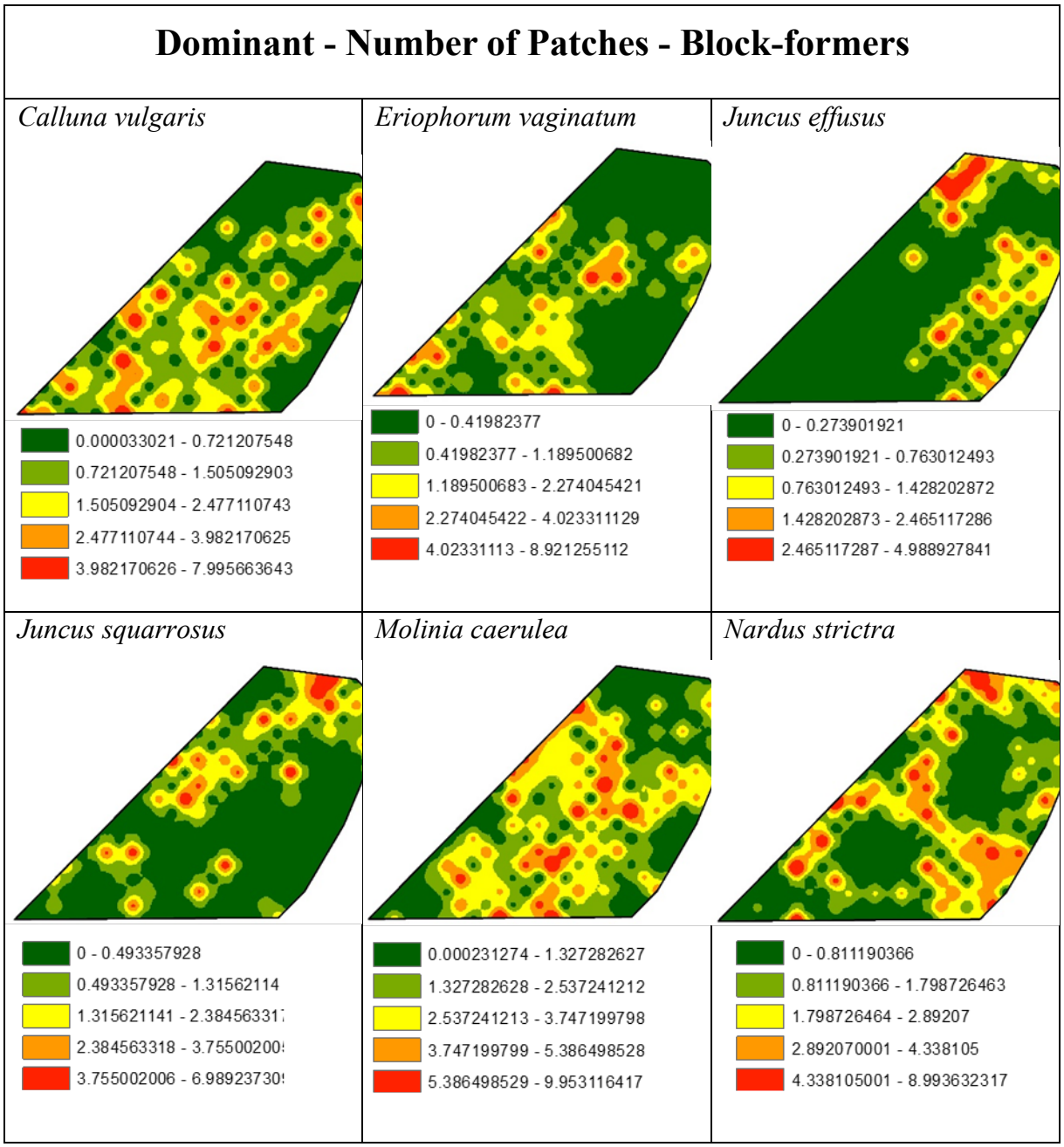


Fig. S2. 2 Interpolation maps of predicted number of patches of selected dominant block-forming vegetation species across the Ashtrees Dipper

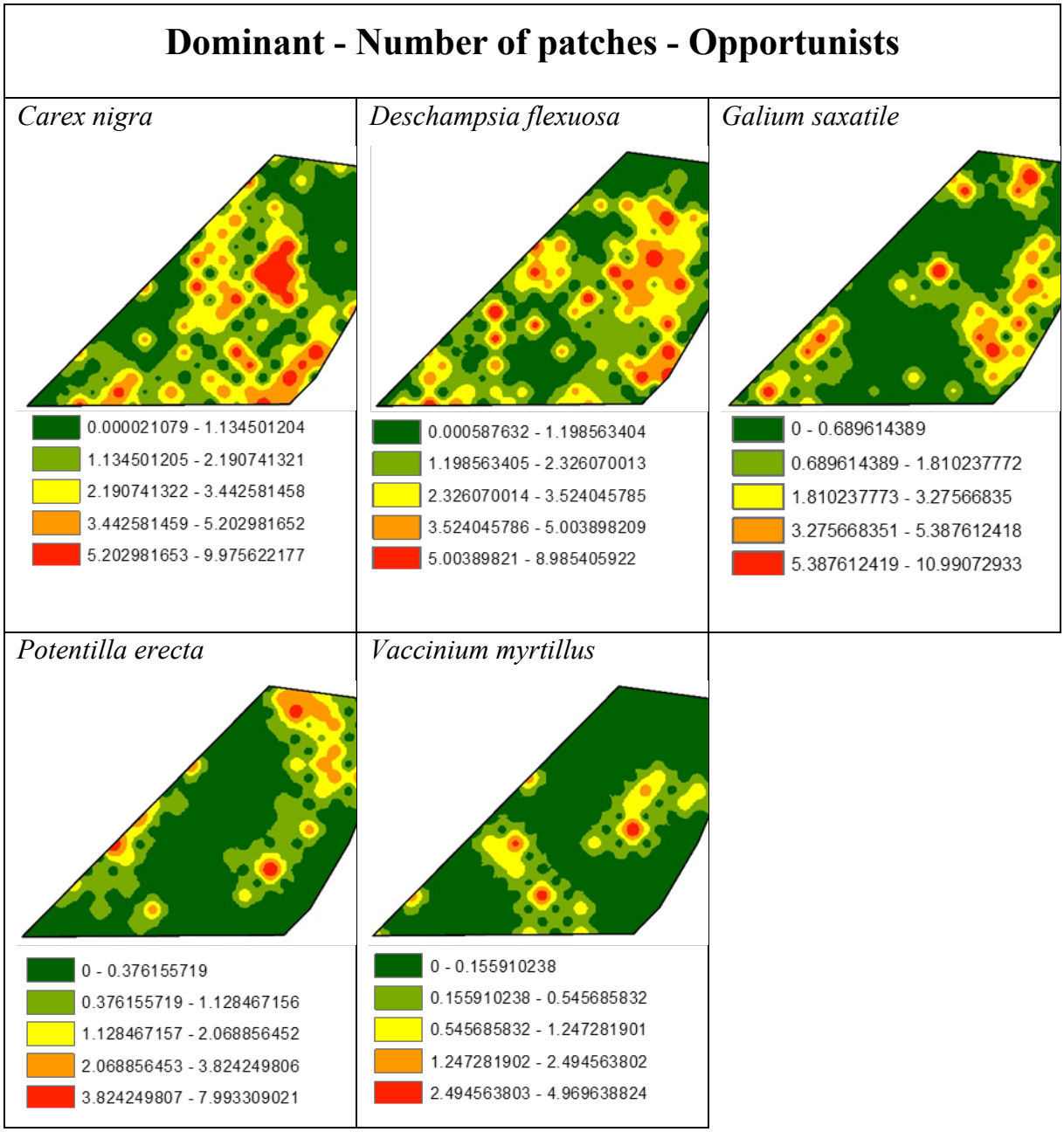


Fig. S2. 3 Interpolation maps of predicted number of patches of selected dominant opportunistic vegetation species across the Ashtrees Dipper

Appendix 1.3. *Area*

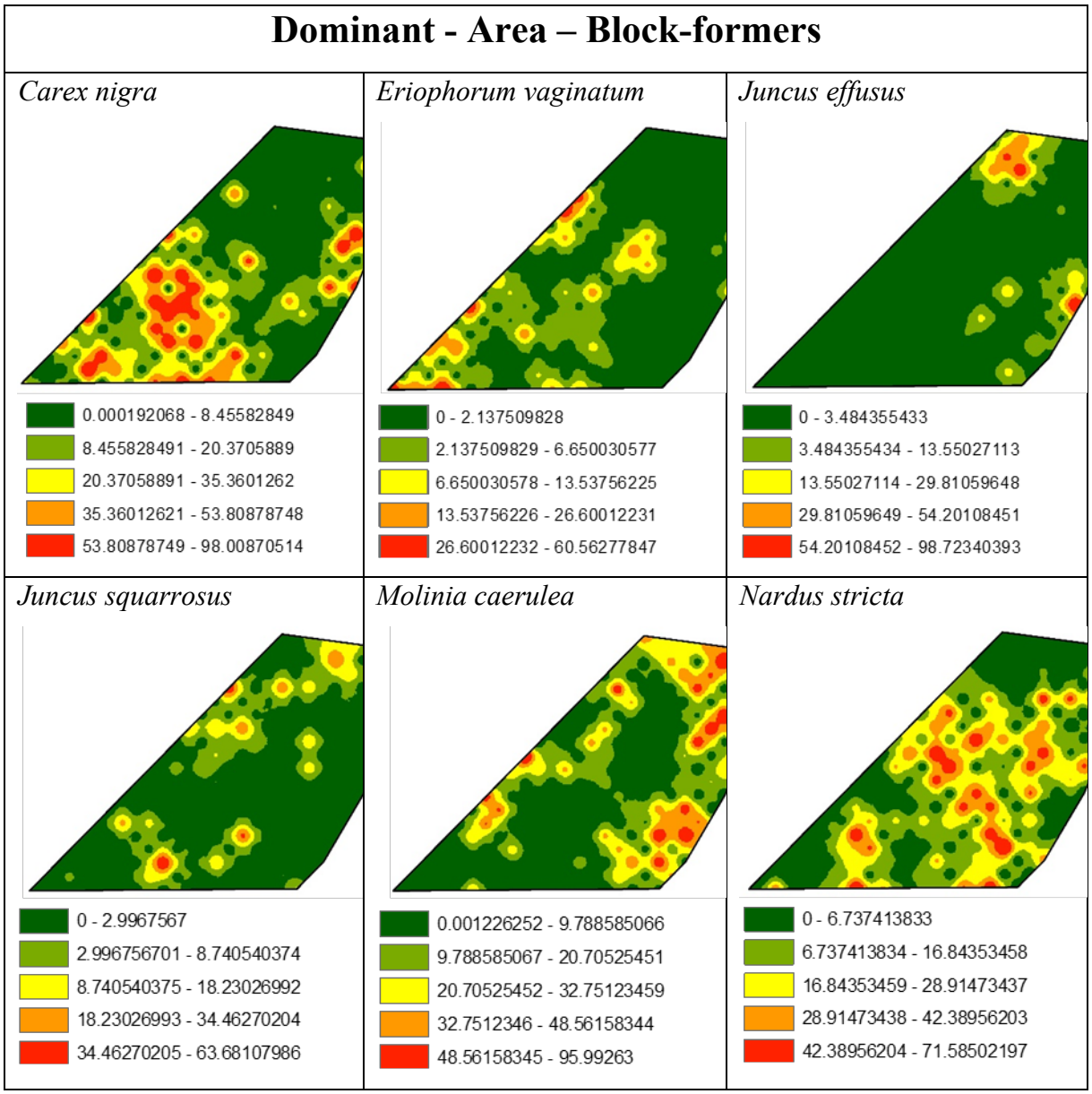


Fig. S2. 4 Interpolation maps of predicted area of selected dominant block-forming vegetation species across the Ashtrees Dipper

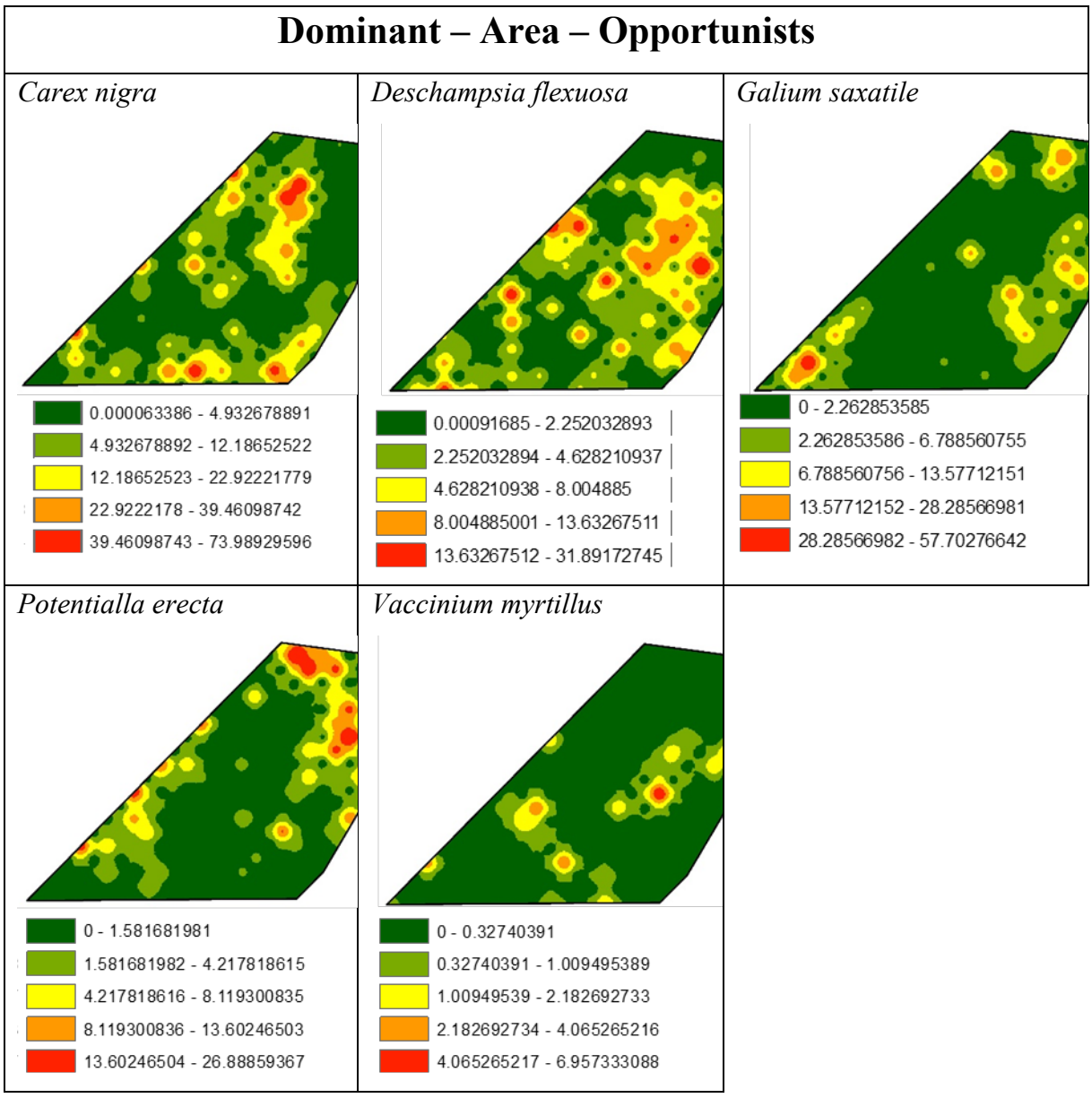


Fig. S2. 5 Interpolation maps of predicted area of selected dominant opportunistic vegetation species across the Ashtrees Dipper

Appendix 1.4. *Shape Index*

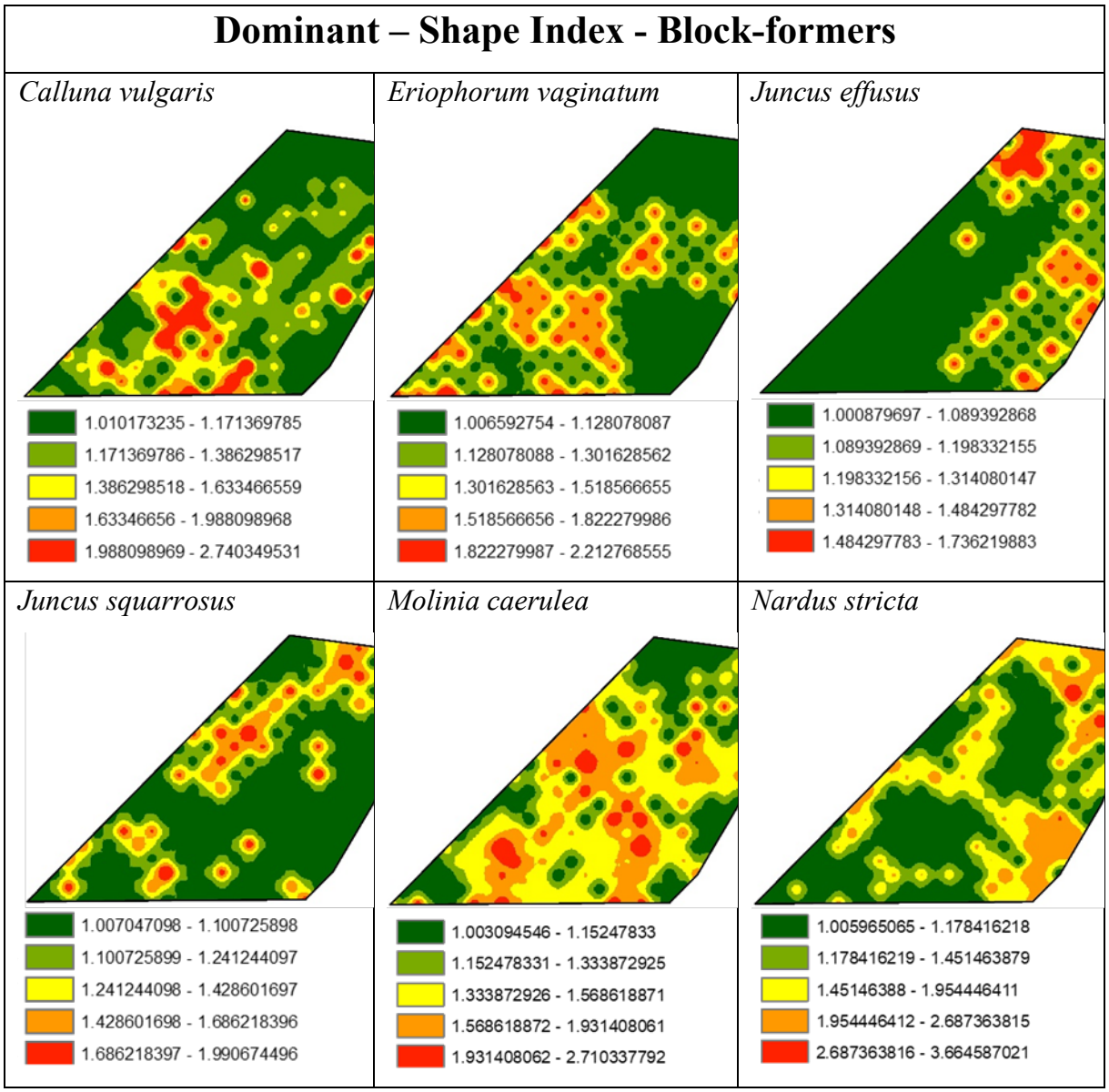


Fig. S2. 6 Interpolation maps of predicted shape index of selected dominant block-forming vegetation species across the Ashtrees Dipper

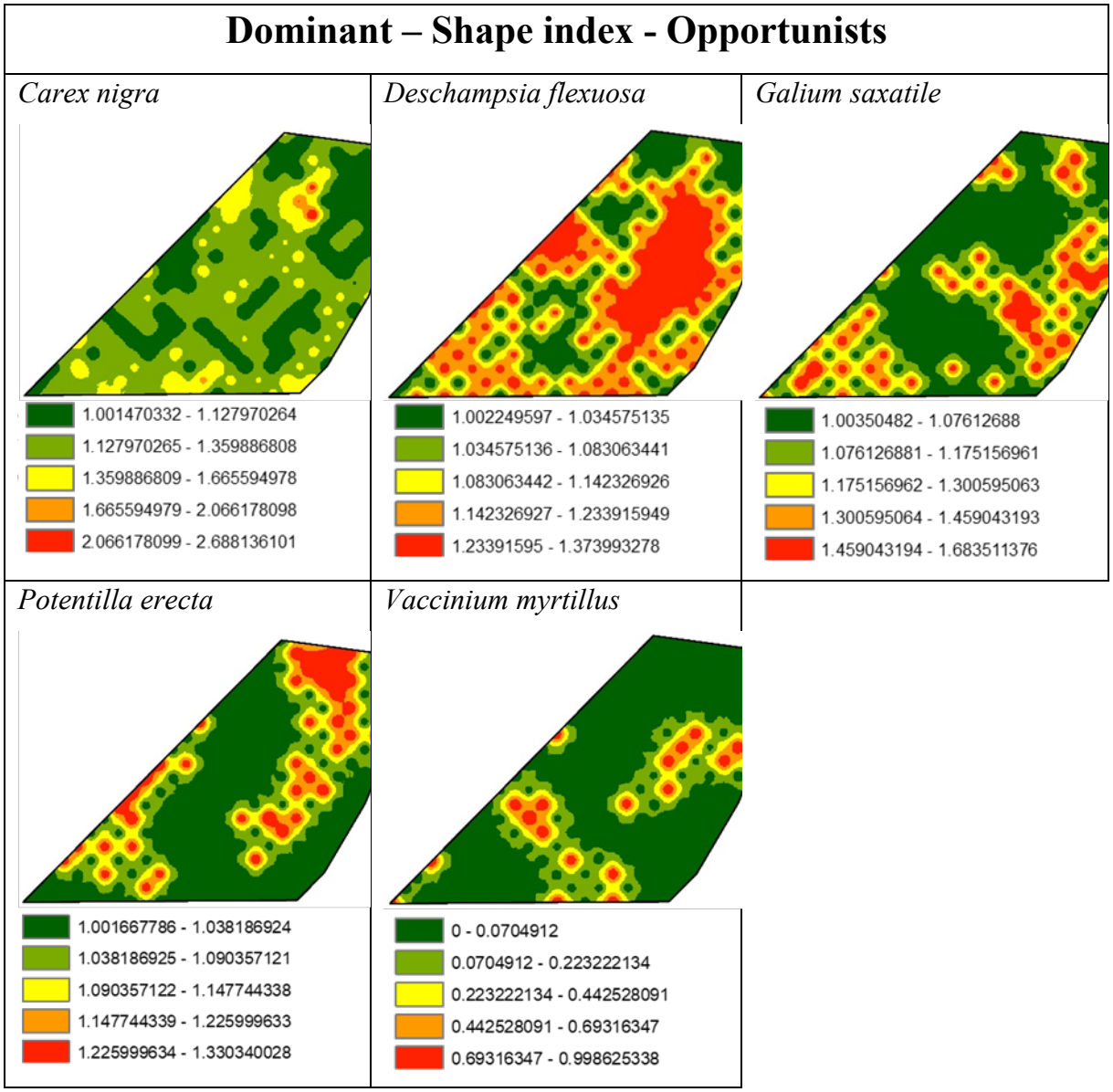


Fig. S2. 7 Interpolation maps of predicted area of selected dominant opportunistic vegetation species across the Ashtrees Dipper

Appendix 2. *UK BAP and Ashtrees NVC Predictions*

Appendix 2.1. UKBAP RF predictions across England and Wales

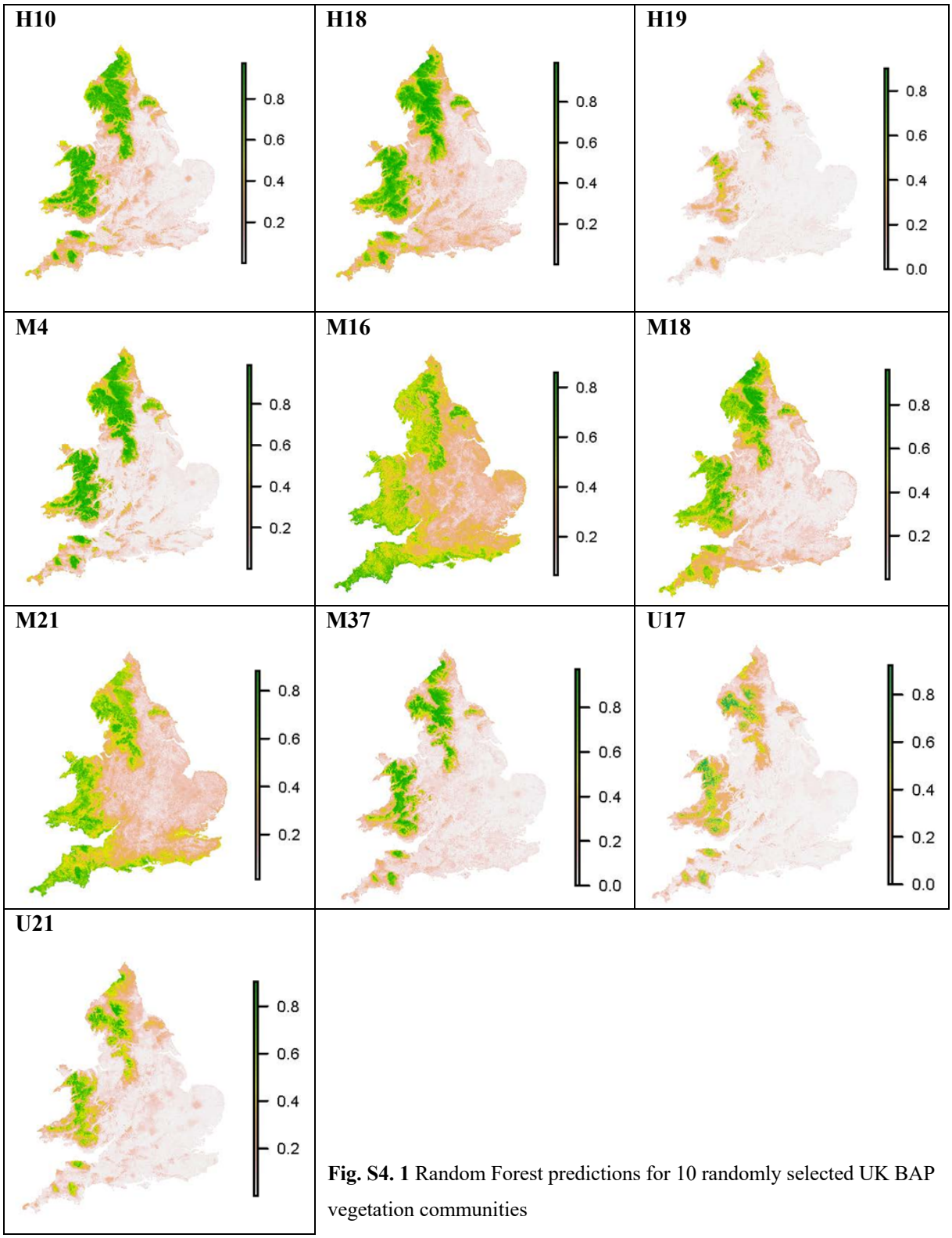


Fig. S4. 1 Random Forest predictions for 10 randomly selected UK BAP vegetation communities

Appendix 2.2. Ashtrees RF predictions across England and Wales

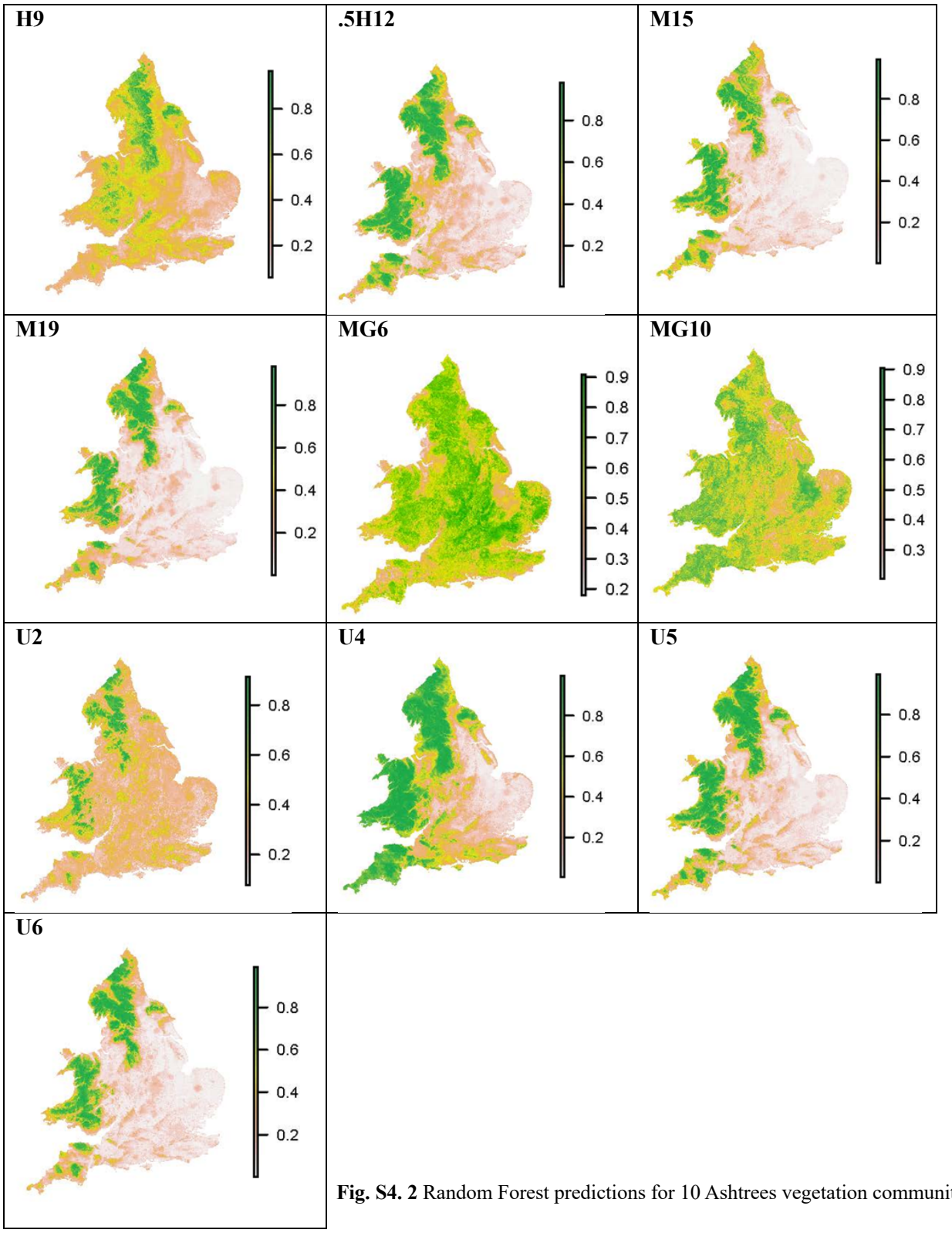
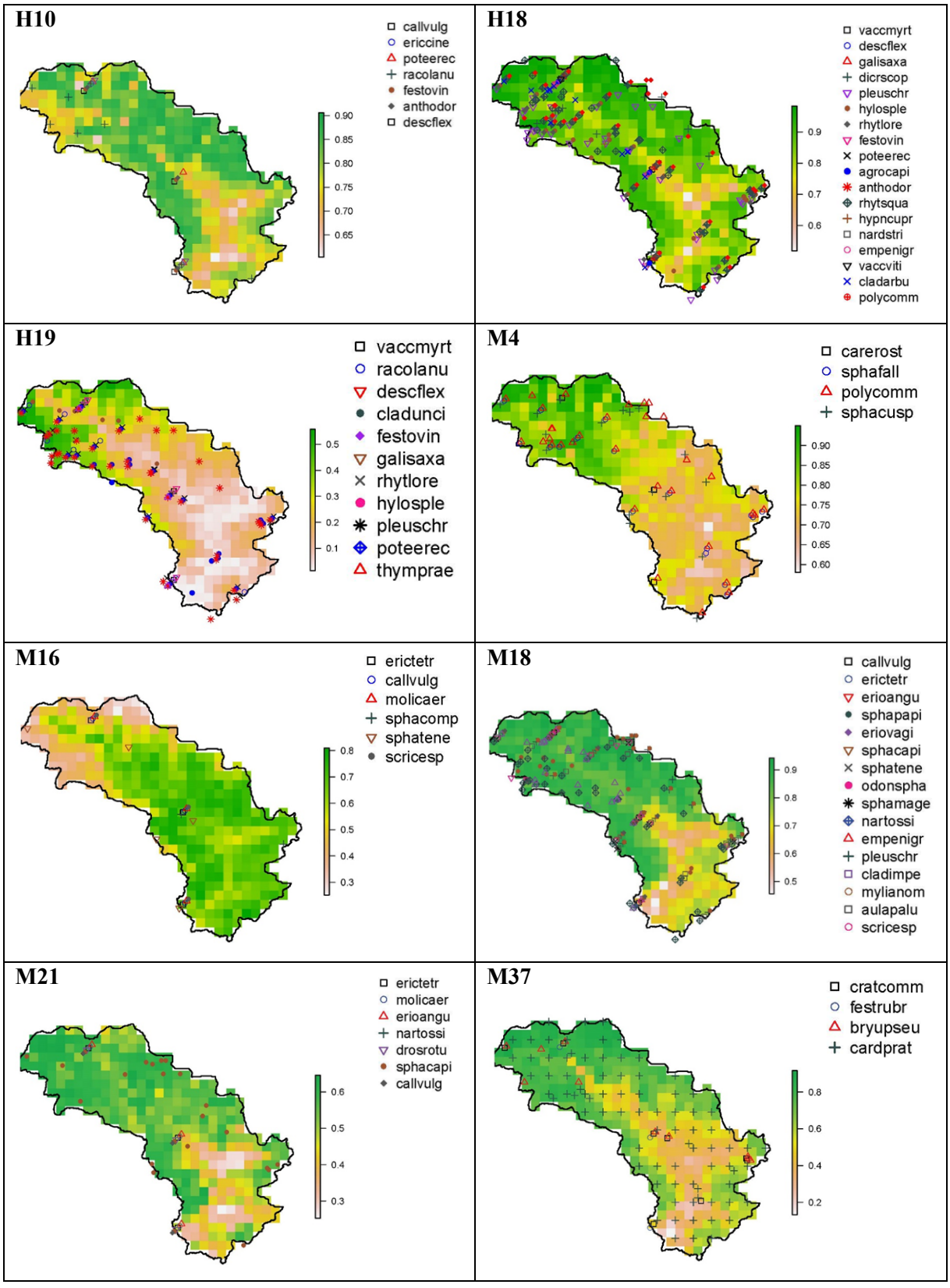


Fig. S4. 2 Random Forest predictions for 10 Ashtrees vegetation communities

Appendix 2.3. UK BAP NBN Atlas Species records in Rede Catchment



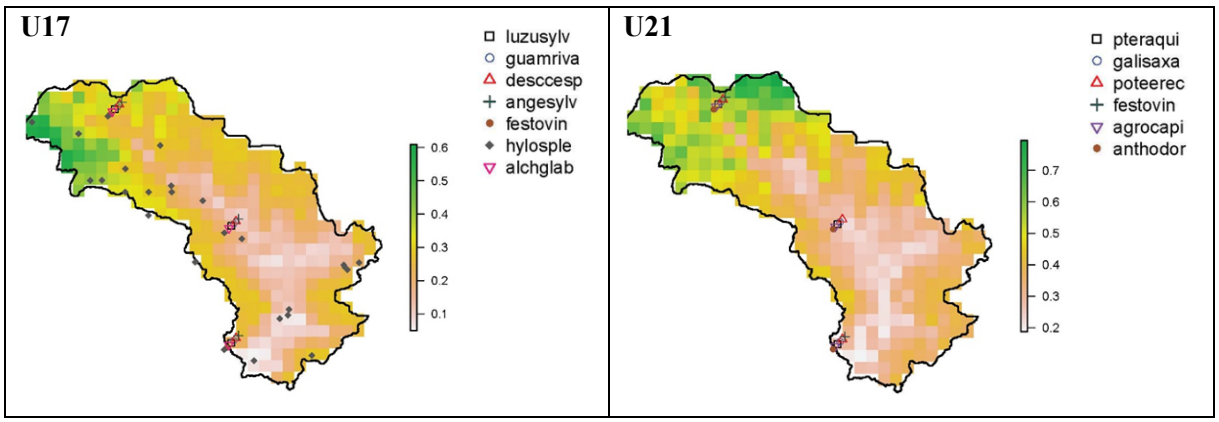
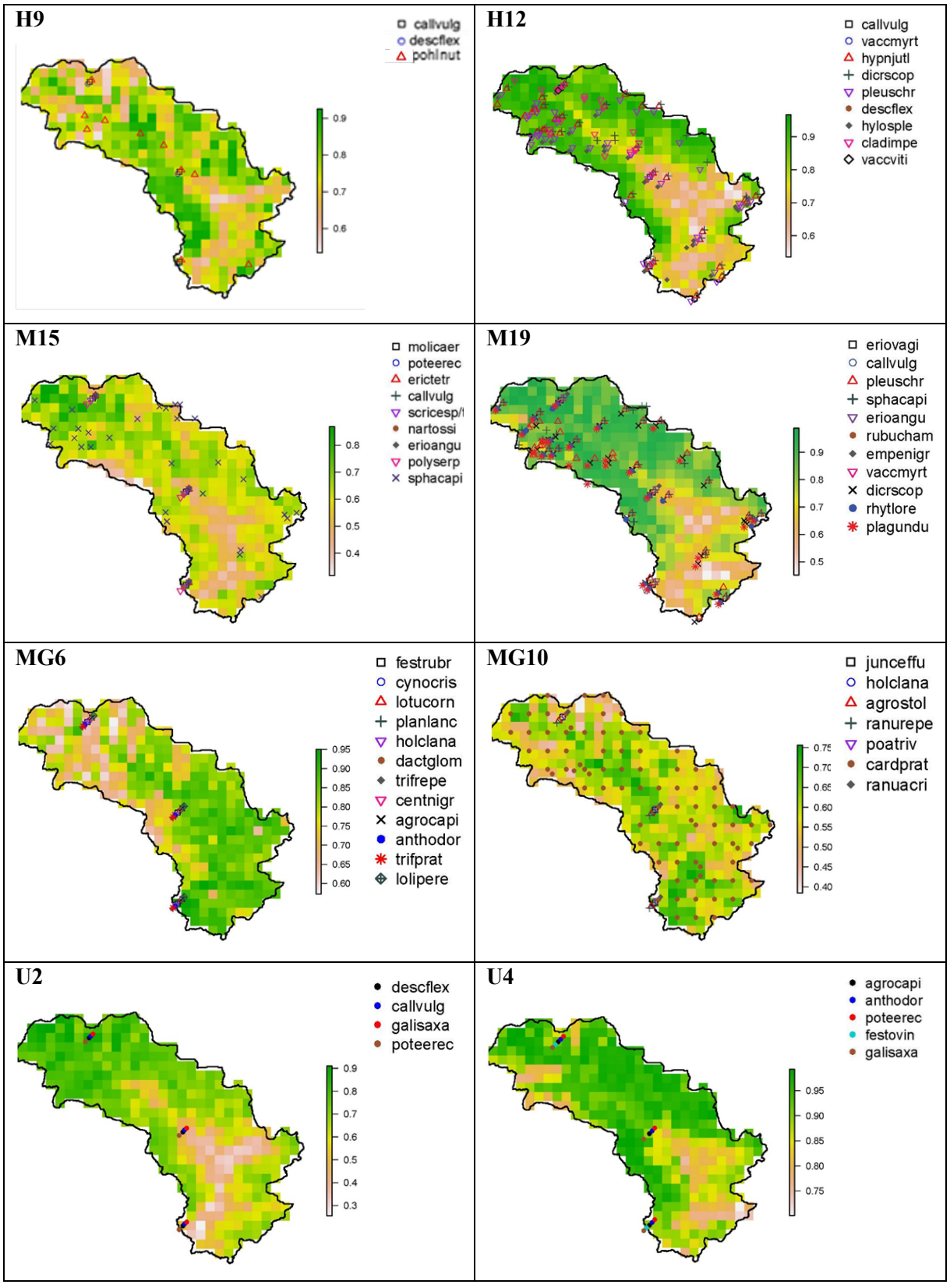


Fig. S4. 3 Comparison of probability of occurrence of the 10 randomly selected UK BAP communities using species data from the NBN Atlas to assess reliability of community predictions.

Appendix 2.4. Ashtrees NBN Atlas Species records in Rede Catchment



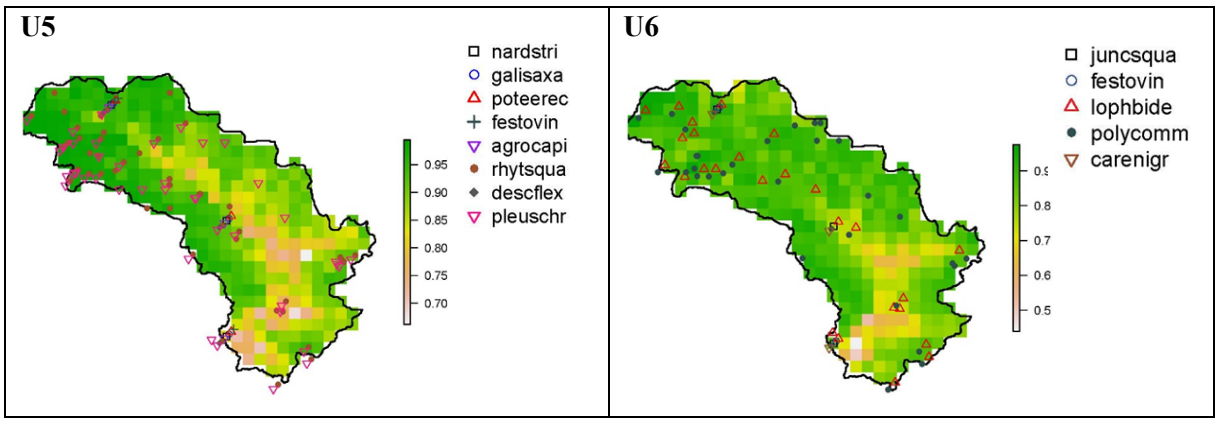


Fig. S4. 4 Comparison of probability of occurrence of the 10 Ashtrees communities using species data from the NBN Atlas to assess reliability of community predictions.

Appendix SUP1. Example *Rscript* Chapter 2

```
rm(list=ls())
library(vegan)
library(dplyr)
library(tidyverse)
library(SDMTools)
library(tidyr)
library(mvabund)
```

Ordination plots for all species at Ashtrees

```
spp_abund <- read.csv("Data/Abundance_2-08-18.csv")
spp_names <- names(spp_abund)
quad_no <- (spp_abund[,1])
spp_abund <- spp_abund[,-1]
```

Principal Component Analysis (pca)

```
abund_ca <- cca(spp_abund)
spp_abund_rda <- rda(spp_abund)
plot(spp_abund_rda, display="sites") # big wedge
plot(spp_abund_rda, display = "species") # big wedge
```

Detrended Correspondance Analysis

```
spp_abund_stand <- decostand(spp_abund, method = 'hellinger')
spp_abund_dca <- decorana(spp_abund_stand)
plot(spp_abund_dca, display = "sites", cex = 0.5)
plot(spp_abund_dca, display = "species", cex = 0.5)
```

Non-Metric Multidimensional Scaling

```
spp_abund_nmds <- metaMDS(spp_abund)
par(pch = 20, cex = 1)
plot(spp_abund_nmds, type = "n", main = "% cover nMDS")
plot(spp_abund_nmds, type = "n", main = "% cover species")
text(spp_abund_nmds, cex = 0.6, display = "species", offset = 0.5)
```

Ordination for dominant species (patch area) and sub dominant species

(This was also performed on subdominant species)

```
area_dom <- read.csv("Results/Patch_Stats/Ordination/Area_patches_spp/Area
_matrix/Area_spp_dominant_4-08-19.csv")
area_dom[is.na(area_dom)] <- 0
```

Using NMDS

```
area_dom.nmds <- metaMDS(area_dom)
plot(area_dom.nmds, type = "n", main = "NMDS Dominant Species")
text(area_dom.nmds, display = "species", cex = 0.7)
points(area_dom.nmds, display = "species", pch = 20, cex = 0.7)
plot(area_dom.nmds, type = "n", main = "NMDS Dominant Quadrats")
text(area_dom.nmds, display = "sites", cex = 0.7)
points(area_dom.nmds, display = "sites", pch = 20)
```

Procrustes rotation dominant vs % Ashtrees

```
area_dom_proc <- procrustes(scores(spp_abund_nmds), scores(area_dom.nmds),
      scale = TRUE, scores = “sites”)
plot(area_dom_proc, cex = 0.6, pch = 20, main = “Procrustes Rotation”)
plot(area_dom_proc, dispay = “target”)
proc_test_dom <- protest(scores(spp_abund_nmds), scores(area_dom.nmds),
      scores = “sites”)
 proc_test_dom <- protest(scores(spp_abund_nmds), scores(area_dom.nmds),
      scores = “species”)
capture.output(proc_test_dom, file =
      “Results/Patch_Stats/Ordination/Procrustes_rotation.csv”)
proc_resid <- residuals(area_dom_proc) # residuals
plot(proc_resid, pch = 20, cex = 0.6, main = “Residuals”) # plot residuals
proc_pred_dom <- predict(area_dom_proc, newdata =
      scores(area_dom.nmds$points), truemean = TRUE) # predict newdata on r
eal scores
plot(proc_pred, type = “n”, main = “Procrustes Prediction”) # plot predict
ed
values
points(proc_pred_dom, cex = 0.6, pch = 20, col = “red”) # plotting predict
ed values following procrustes
```

Procrustes rotation for dominant vs subdominant

```
area_dom_subdom_proc <- procrustes(scores(area_dom.nmds), scores(area_subd
om.nmds), scale = TRUE, scores = “sites”)
summary(area_dom_subdom_proc)
plot(area_dom_subdom_proc, cex = 0.6, pch = 20, main = “Procrustes Rotatio
n”) plot(area_dom_subdom_proc, dispay = “rotation”)
capture.output(summary(area_dom_subdom_proc), file =
      “Results/Patch_Stats/Ordination/Procrustes_rotation_dom-subdom_summar
y.csv”)
```

Patch Statistics on Individual Quadrats

```
Dominant <- data.frame(read.csv(“Data/Dominant-sub_New_Versions/Dominant_2
3-7-18_spaced.csv”, header = FALSE)) # main datafram (all inputs)
colnames(Dominant) <- c(“quad_no”, “Row”, 10, 20, 30, 40, 50, 60, 70, 80,
90, 100) Dominant[1,] = "" ## add empty line to streamline with other quad
rat rows Dominant[1838,] = "" ## add line after last quadrat to streamline
with other quadrat rows
quad1_to_167 <- Dominant
quad_dom_mat_all <- as.matrix(quad1_to_167[,-1])
quad_dom_tab_all <- table(quad_dom_mat_all[,-1]) # gives count of each spe
cies in the quadrat quad_dom_tab_all
df_dom <- data.frame(quad_dom_tab_all)
quad_lng <- quad1_to_167 %>% gather(10:100, -quad_no, key=“Col”,
      value=“Species”) # drops variables since they are not identical acrio
      ss mewasures/This is always used before filtering by species name
summary(quad_lng)
quad_lng <- mutate(quad_lng, Row = as.factor(Row), Col = as.factor(Col)) #
This is always used before filtering by species name
levels(quad_lng$Col) <- c(“10”, “20”, “30”, “40”, “50”, “60”, “70”, “80”,
      “90”, “100”)  # This is always used before filtering by species name
```

Generating a loop to read through all quadrats

```
spp_in_quad <- names(quad_dom_tab)
spat_all <- NULL
for (spp_name in seq_along(spp_in_quad)) {
      spp_text <- names(quad_dom_tab)[spp_name]
      quad_spp_lng <- quad_lng %>% filter(Species == spp_text)
      quad_spp_wde <- quad_spp_lng %>% reshape2::dcast(Row ~ Col, drop = FA
LSE, fill = 0) #Recode text to number
      quad_spp_wde[,-1] <- ifelse (quad_spp_wde[,-1]!="0",1,0)
      quad_spp_mat <- as.matrix(quad_spp_wde[,-1])
      quad_ccl = ConnCompLabel(quad_spp_mat)
      print(spp_text)
      print(quad_ccl)
      image(t(quad_ccl[10:1,]),col=c('white',rainbow(length(unique(quad_ccl
))
      -1)))
      grid(nx, ny, col = "grey", lty = 1) # added post
      print(PatchStat(quad_ccl))

# Store the spatial stats
# Need to duplicate spp text names in final output table
spp_duplicate <- rep(spp_text, max(quad_ccl)+1)
patch_stats <- PatchStat(quad_ccl)
patch_stats <- cbind(spp_duplicate, patch_stats)
spat_all <- rbind(spat_all, patch_stats) #readline() }
write.csv(spat_all, file = "Results/Patch_Stats/Quad_dom_1-167_spaced.csv"
)
```

PATCH STATS: Number of for chosen species per quadrats

(this was repeated for all dominant and subdominant block-forming and opportunistic species)

Calluna Vulgaris

```
callvulg_dom <- subset(patch_stats_dom, spp_duplicate == "Callvulg")
callvulg_dom <- subset(callvulg_dom, patchID > 0)
patchno_callvulg_dom<- aggregate(callvulg_dom$patchID, by = list(quad_no =
callvulg_dom$quad_no), FUN = max) # number of patches per species write.cs
v(patchno_callvulg_dom, file =
      "Results/Patch_Stats/patchno/callvulg.csv")
```

Nardus Stricta

```
nardstri_dom <- subset(patch_stats_dom, spp_duplicate == "Nardstri")
nardstri_dom <- subset(nardstri_dom, patchID > 0)
patchno_nardstri_dom<- aggregate(nardstri_dom$patchID, by = list(quad_no =
nardstri_dom$quad_no), FUN = max) # number of patches per species write.cs
v(patchno_nardstri_dom, file =
      "Results/Patch_Stats/patchno/nardstri.csv")
```

Number of Patches: ManyGLM Analysis

(This was repeated for Area and Shape Index of patches of dominant and subdominant species)

Dominant species

```
patchno_dom_spp <- read.csv("Results/Patch_Stats/Ordination/Patchno_spp/Patchno_matrix/
      Patchno_matrix_dom_12-08-19.csv")

patchno_dom_spp[is.na(patchno_dom_spp)] <- 0
env_var <- data.frame(read.csv("Data/Biotic_data/ashtrees_env_variables_03-
      08-18.csv"))

patchno_dom_mvabund <- mvabund(patchno_dom_spp)
      plot.mvformula(log(patchno_dom_mvabund+1) ~ exp(env_var$soil_pH), main="Dominant Patchno vs Soil pH", xlab="% Soil pH - Log Scale", ylab="
      Abundance [log scale]", overall.main="Species Abundance vs soil pH",
      fg="grey", las=1, scale.lab="ss",t.lab="o", mfrow=c(4,3),log="x")

# Produces a range of plots for visualising multivariate abundance data and its relationship to environmental variables
boxplot(patchno_dom_mvabund, horizontal = TRUE, las = 2, main = "Number of Patches") # boxplot
meanvar.plot(patchno_dom_mvabund) # check mean variance
patchno_dom_glm_negbinomial <- manyglm(patchno_dom_mvabund ~ soil_pH + slope
      + pct_water + Altitude + length_10m + distance_10m + length_25m + distance_25m + length_35m + distance_35m, data = env_var, family = "negative.binomial") # does not give a particular shape; will use this

plot(patchno_dom_glm_negbinomial)
summary(patchno_sub_glm_negbinomial, resamp="monte.carlo", test="wald", nBoot=300) # summary method for class "manyglm"

anova_glm_dom_patchno <- anova.manyglm(patchno_dom_glm_negbinomial, show.time
      = "all") capture.output(anova_glm_dom_patchno, file = "Results/DOMINANT/anova_glm_dom_patchno.doc")

anova_ind_dom_patchno <- anova(patchno_dom_glm_negbinomial, p.uni="adjusted",
      show.time = "all")
capture.output(anova_ind_dom_patchno, file = "Results/DOMINANT/patchno_anova_glm_dom_ind.csv")
```

Appendix SUP2. Example *Rscript* Chapter 3

```
rm(list=ls())
library(vegan)
library(dplyr)
library(tidyverse)
```

Using twinspanR for clustering

```
ash.twi.whittaker8 <- twinspan(ashtrees.mod, modif = TRUE, clusters
=
      8, diss = 'whittaker')
k.whittaker8 <- cut(ash.twi.whittaker8)
k.whittaker.df <- data.frame(Quad = rownames(ashtrees.mod), Cluster
=
      k.whittaker)
ashtrees.NMDS # use NMDS results instead of decorana
par(mfrow = c(1,2), pch = 20, cex = 0.9)
ordiplot(ashtrees.NMDS, type = 'n', display = 'si')
points(ashtrees.NMDS, col = k.whittaker8)
for (i in c(1:9)) ordihull (ashtrees.NMDS, groups = k.whittaker8,
      show.group = i, col = i, draw = 'polygon', label = TRUE)
text(ashtrees.NMDS, display="sites", cex=0.6, pos = 2)

ordiplot(ashtrees.NMDS, type = 'n', display = 'si', main = 'Modified
TWINSPAN\n Method=Whittaker9')
points(ashtrees.NMDS, col = k.whittaker9)
for (i in c(1:9)) ordihull (ashtrees.NMDS, groups = k.whittaker9,
show.group = i, col = i, draw = 'polygon', label = TRUE)
```

Generating Random Quadrats from Literature

(Using 4b sub-community as an example)

```
## U4b
U4b <- read.csv("Rodwell_Literature/U4b.csv")
set.seed(121)
sample.no <- 25
FINAL.U4b=NULL

for(quad_no in 1:sample.no){
  print(quad_no)
  No.spp <- as.integer(runif(1, min=1, max=19))
  U4b.pseudo <- sample_n(U4b, size=No.spp, replace=FALSE)
  U4b.pseudo <- as.data.frame(U4b.pseudo)
  TEMP.output <- data.frame(cbind(quad_no, U4b.pseudo))
  FINAL.U4b <- rbind(FINAL.U4b, TEMP.output)
}
write.csv(FINAL.U4b, file = "Rodwell_Literature/U4b_pseudo.csv")
```

Plotting the results in ordination with centroids

```
# Using DECORANA to be in line with pseudo quadrat literature
analysis
```

```
#ashtrees.dca.xy <- data.frame(scores(ashtrees.dca, display =
"sites"))

## Adding centroids to scatter plot/ordination plot and saving to
pdf
#x.ash <- ashtrees.dca.xy$DCA1
#y.ash <- ashtrees.dca.xy$DCA2

class.ash

# class <- data.frame(class=class[-189:-nrow(class),])
df.ash <- data.frame(class.ash, x.ash, y.ash)
colnames(df.ash) <- c("class", "x", "y")

centroids.ash <- merge(df.ash, aggregate(cbind(mean.x=x,
mean.y=y)~class, df.ash, mean), by="class")
ggplot(centroids.ash, aes(x, y, color = factor(class))) +
geom_point(size=3) + geom_point(aes(x=mean.x, y=mean.y), size = 5) +
  geom_segment(aes(x=mean.x, y=mean.y, xend=x, yend=y))

Plot.dca.ash <- ggplot(centroids.ash, aes(x, y, color =
factor(class))) + geom_point(size=0.8) + geom_point(aes(x=mean.x,
y=mean.y), size = 3) +
  geom_segment(aes(x=mean.x, y=mean.y, xend=x, yend=y))

plot.1.edit <- Plot.dca.ash + theme(panel.grid.major =
element_blank(), panel.grid.minor = element_blank(), # removes
major/minor gridlines
                                    panel.background =
element_blank(), # removes grey background
                                    axis.line = element_line(colour
= "black"), # colours axes to black
# legend.position = "none", # removes legends
plot.title = element_text(hjust = 0.5), # centres graph title
legend.background = element_blank(), # removes legend background
legend.key = element_blank()) # removes boxes behid points in legend
```

Plotting vertical and horizontal error bars for quadrats/communities for pseudoquadrats generated from literature

```
centroids.2.ash <- aggregate(cbind(x,y)~class, df.ash, mean) #
aggregating dataframe
f <- function(z)sd(z)/sqrt(length(z)) # Setting function to
calculate standard error
se.ash <- aggregate(cbind(se.x=x, se.y=y)~class, df.ash, f) #
calculating standard error
cent.2.ash <- merge(centroids.2.ash, se.ash, by="class") # adding
standard error column to centroids
plot.2.ash <- ggplot(cent.2.ash, aes(x,y,color = factor(class))) + #
sets plot aesthetics of x, y and centroids
```

```
  scale_colour_discrete(l=0) + # sets centroids and errorbars to 1
colour (black by default)
  geom_point(data=centroids.2.ash, size = 1) + # sets aesthetics of
point plots
  geom_errorbar(data=cent.2.ash, aes(ymin=y-se.y, ymax=y+se.y),
width = 0.01) + # sets vertical error bars
  geom_errorbarh(data=cent.2.ash, aes(xmin=x-se.x, xmax=x+se.x),
height = 0.01)  # sets horizontal error bars

# Editing the graph (remove grey background, grid lines and adding
axis lines)
Plot.2 <- plot.2.ash + theme(panel.grid.major = element_blank(),
panel.grid.minor = element_blank(), # removes major/minor gridlines
panel.background = element_blank(), # removes grey background
axis.line = element_line(colour = "black"), # colours axes to black
legend.position = "none", # removes legends
plot.title = element_text(hjust = 0.5)) # centres graph title
# legend.background = element_blank(), # removes legend background
# legend.key = element_blank()) # removes boxes behid points in
legend
```

Predicting where ashtrees points fall in literature DCA ordination

```
plot.2.dca <- ggplot(cent.add.2, aes(x,y,color = factor(class))) + #
sets plot aesthetics of x, y and centroids
  scale_colour_discrete(l=0) + # sets centroids and errorbars to 1
colour (black by default)
  geom_point(data=cent.add.2, size = 1) + # sets aesthetics of point
plots
  geom_errorbar(data=cent.add.2, aes(ymin=y-se.y, ymax=y+se.y),
width = 0.05) + # sets vertical error bars
  geom_errorbarh(data=cent.add.2, aes(xmin=x-se.x, xmax=x+se.x),
height = 0.05)  # sets horizontal error bars

# Editing the graph (remove grey background, grid lines and adding
axis lines)
Plot.2.dca.edit <- plot.2.dca + theme(panel.grid.major =
element_blank(), panel.grid.minor = element_blank(), # removes
major/minor gridlines
panel.background = element_blank(), # removes grey background
axis.line = element_line(colour = "black"), # colours axes to black
legend.position = "none", # removes legends
plot.title = element_text(hjust = 0.5)) # centres graph title
##, legend.background = element_blank(), # removes legend background
## legend.key = element_blank()) # removes boxes behid points in
legend

print(Plot.2.dca.edit + ggtitle("Pseudoquadrats for literature
species found in plant communities at Ashtrees")
     + labs(x="DCA1", y="DCA2", color = "Community") +
geom_point(data=cent.add.2[1:2,], aes(x=x, y=y), colour="red")
```

```
     + geom_errorbar(data=cent.add.2[1:2,], aes(ymin=y-se.y,
ymax=y+se.y), colour = "red", width = 0.05)
     + geom_errorbarh(data=cent.add.2[1:2,], aes(xmin=x-se.x,
xmax=x+se.x), colour = "red", height = 0.05)
      + geom_text_repel(data=cent.add.2, aes(x, y, label=class),
                        nudge_x=0.13, nudge_y=0.13, size = 3))


predict.communities <- as.data.frame(predict(quads.dca,
ashtrees.mod, type = "sites"))

Plot3 <- Plot.2.dca.edit + geom_point(data = predict.communities,
aes(x=DCA1, y=DCA2, colour="black"))

#print(Plot.2.dca.edit + ggtitle("Pseudoquadrats for Ashtrees
species vs. Ashtrees pseudoquadrats")
#      + geom_point()
#      + geom_text(data = predict.communities, aes(x=DCA1, y=DCA2,
label=rownames(predict.communities)), colour="red", size = 3)
#      + labs(x="DCA1", y="DCA2", color = "Community") +
#        geom_text_repel(data=plot.2.centroids, aes(x, y,
label=class),
#                        nudge_x=0.14, nudge_y=0.09, size = 3))
write.csv(predict.communities,
file="Ashtrees/Raw/Prediction/Ash_Lit.csv")
```

Generate confusion matrices

```
LIT.MAVIS <-read.csv("Ashtrees/Predictions/LIT_vs_MAVIS.csv",
row.names = 1)
LIT.MAVIS[is.na(LIT.MAVIS)] = 0
LIT.MAVIS <- as.matrix(LIT.MAVIS)
LIT.MAVIS <- as.table(LIT.MAVIS)
LIT.conf <- confusionMatrix(LIT.MAVIS, positive = NULL, prevalance =
NULL, mode = "sens_spec")

ASH.MAVIS <- read.csv("Ashtrees/Predictions/ASH_vs_MAVIS.csv",
row.names = 1)
ASH.MAVIS[is.na(ASH.MAVIS)] = 0
ASH.MAVIS <- as.matrix(ASH.MAVIS)
ASH.MAVIS <- as.table(ASH.MAVIS)
ASH.conf <- confusionMatrix(ASH.MAVIS, positive = NULL, prevalance =
NULL, mode = "sens_spec")
```

Appendix SUP3. Example *Rscript* Chapter 4

```
rm(list = ls())
graphics.off()

install.packages("remotes")
remotes::install_github("johnbaums/rmaxent")

library(rJava)
library(dismo)
library(kernlab)
library(rmaxent)
library(maxlike)
library(randomForest)
library(LSRS)
library(sf) # installed
library(tmap)
library(lwgeom)
library(gdtools) # installedl requiresd pkg-config and cairo from brew
library(sp)
library(rgdal)
library(lattice)
library(latticeExtra)
library(mapview) # Roy install
library(rgeos)
library(reshape2)
library(plyr)
library(dplyr)
library(raster)
library(RColorBrewer)
library(spacetime)
library(stringr)
library(gdalUtils)
library(gstat)
library(devtools)
library(vegan)
library(stars)
library(sdm)
library(SDMTools)
install.packages(c("rgdal", "lattice", "latticeExtra", "mapview", "reshape
2", "plyr", "dplyr", "raster", "RColorBrewer", "spacetime", "stringr", "gd
alUtils",
                   "gstat", "devtools", "vegan", "stars", "sdm", "SDMTools
", "dismo"))
#### START ####
crs_UK <- CRS("+init=epsg:27700") # set CRS; we will be using British Nati
onal Grid EPSG:27700
crs_latlong <- CRS("+init=epsg:4326")

#### Using GB outlines from different sources ####

#### UK and Wales outlines ####

England_outline <- read_sf("DATA/GIS_data/gb-outlines/england.shp")
Wales_outline <- read_sf("DATA/GIS_data/gb-outlines/wales.shp")
Wales_outline <- Wales_outline[,-1]
```

```
england_wales <- st_union(England_outline, Wales_outline) # England and wa
les for use with soil maps
england_wales <- st_transform(england_wales, 27700)


#### LANDCOVER 2015 1KM ####
LCM2015_1KM_aggr_gb <- raster("DATA/GIS_data/LANDCOVER/lcm-2015-1km_250393
3/aggregate_class/lcm2015_gb_1km_dominant_aggregate_class.tif") # loading
the 25m raster but seems to have a problem with colouration
LCM2015_1KM_aggr_gb
plot(LCM2015_1KM_aggr_gb)
colors_LCM2015_1KM_aggr_gb <- c("white", "red", "darkgreen", "brown", "gre
en", "darkolivegreen", "plum1", "darkblue", "royalblue3", "lightyellow", "
grey") # set colors according to landcover map
breaks_LCM2015_1KM_aggr_gb <- c(0:10) #set breaks
plot_LCM2015_1KM_aggr_gb <- plot(LCM2015_1KM_aggr_gb, col = (colors_LCM201
5_1KM_aggr_gb), breaks = breaks_LCM2015_1KM_aggr_gb, legend = FALSE, axes
= FALSE, box = FALSE)
par(xpd = TRUE)
legend_LCM2015_1KM_aggr_gb <- c("Unclassified", "Broadleaved woodland", "C
oniferous woodland", "Arable", "Improved grasslands",
                     "Semi-natural grasslands", "Mountain, heath and bog","
Saltwater", "Freshwater", "Coastal", "Built-up areas and gardens")
LCM2015_classifications <- data.frame(legend_LCM2015_1KM_aggr_gb) # classi
fcation as dataframe
legend("topright", legend = legend_LCM2015_1KM_aggr_gb, cex = 0.5, fill =
colors_LCM2015_1KM_aggr_gb, bty = "n") # set legend; bty removes legend bo
rder

## Soil Data ###
soil_uk_NATMAP1000 <- st_read("DATA/GIS_data/SOIL/SOIL/Spatial Soil Data/N
ATMAP1000.shp")
soil_uk_NATMAP1000
soil_uk_NATMAPsoilscapes <- st_read("DATA/GIS_data/SOIL/SOIL/Spatial Soil
Data/NATMAPsoilscapes.shp")
soil_uk_NATMAPsoilscapes
soilscape_types_table <- data.frame(table(soil_uk_NATMAPsoilscapes$SOILSCA
PE))
soilscape_types <- as.character(soilscape_types_table$Var1)

soil_england_wales <- st_intersection(soil_uk_NATMAPsoilscapes, england_wa
les)
st_write(soil_england_wales, dsn = "DATA/GIS_data/SOIL/SOIL/soil_england_w
ales.shp")
soil_england_wales <- read_sf("DATA/GIS_data/SOIL/SOIL/soil_england_wales.
shp")
plot(st_geometry(soil_england_wales))
soil_raster <- raster(soil_england_wales)

## merge LANDSAT ####
UK_rasters <- merge(reproject_brick_201023, reproject_brick_201024, reproj
ect_brick_201025, -reproject_brick_20190130-, reproject_brick_20190202,
                    -reproject_brick_20190130_2-, reproject_brick_20190215
, reproject_brick_202024, reproject_brick_202025, reproject_brick_203024,
reproject_brick_203025, reproject_brick_20180517, reproject_brick_20190128
,
```

```
                    reproject_brick_20180517, reproject_brick_20190128_2,
reproject_brick_20180524,
                    reproject_brick_20190204, reproject_brick_205022, repr
oject_brick_205023, reproject_brick_205024,
                    reproject_brick_206021, reproject_brick_206022, reproj
ect_brick_207021, tolerance = 0.5, filename = "DATA/GIS_data/LANDSAT/Merge
d/UK_merged.tif", overlap = TRUE, ext = NULL)

UK_20123_201024_merge <- merge(reproject_brick_201023, reproject_brick_201
024, tolerance = 0.5, filename = "DATA/GIS_data/LANDSAT/Merged/201023_2010
24.tif",
                                           overlap = TRUE, ext = NULL)
UK_202024_202025_merge <- merge(reproject_brick_202024, reproject_brick_20
2025, tolerance = 0.5, filename = "DATA/GIS_data/LANDSAT/Merged/202022_202
023.tif",
                                overlap = TRUE, ext = NULL)
UK_2019202_2019215 <- merge(mask_20190202, mask_2019215, tolerance = 0.5,
overlap = TRUE, ext = NULL)

England_Wales_spdf <- as(england_wales, 'Spatial') # convert NE_counties t
o SpatialDataframe
crop_201023 <- raster::crop(reproject_brick_201023, England_Wales_spdf)
mask_201023 <- raster::mask(crop_201023, England_Wales_spdf)
crop_201024 <- raster::crop(reproject_brick_201024, England_Wales_spdf)
mask_201024 <- raster::mask(crop_201024, England_Wales_spdf)
crop_201025 <- raster::mask(reproject_brick_201025, England_Wales_spdf)
mask_201025 <- rater::mask(crop_201025, England_Wales_spdf)
crop_2019215 <- raster::crop(reproject_brick_20190215, England_Wales_spdf)
mask_2019215 <- raster::mask(crop_2019215, England_Wales_spdf)
crop_20190202 <- raster::crop(reproject_brick_20190202, England_Wales_spdf
)
mask_20190202 <- raster::mask(crop_20190202, England_Wales_spdf)
crop_202024 <- raster::crop(reproject_brick_202024, England_Wales_spdf)
mask_202024 <- raster::mask(crop_202024, England_Wales_spdf)
crop_202025 <- raster::crop(reproject_brick_202025, England_Wales_spdf)
mask_202025 <- raster::mask(crop_202025, England_Wales_spdf)

plotRGB(mask_201023, r = 6, g = 5, b = 4, maxpixels = 5000000, stretch = N
ULL, scale = max(mask_201023), ext = NULL, colNA = "transparent", add = TR
UE)

UK_rasters_list <- merge(c(extent(reproject_brick_201023), extent(reprojec
t_brick_201024), extent(reproject_brick_201025), extent(reproject_brick_20
190130),
                           extent(reproject_brick_20190202),
                    extent(reproject_brick_20190130_2), extent(reproject_b
rick_20190215), extent(reproject_brick_202024), extent(reproject_brick_202
025),
                    extent(reproject_brick_203024), extent(reproject_brick
_203025), extent(reproject_brick_20180517), extent(reproject_brick_2019012
8),
                    extent(reproject_brick_20180517), extent(reproject_bri
ck_20190128_2), extent(reproject_brick_20180524),
                    extent(reproject_brick_20190204), extent(reproject_bri
ck_205022), extent(reproject_brick_205023), extent(reproject_brick_205024)
,
```

```
                    extent(reproject_brick_206021), extent(reproject_brick
_206022), extent(reproject_brick_207021), overlap = TRUE))

##### TEMP ####
temp1 <- merge(reproject_brick_201023, reproject_brick_201024, tolerance =
0.5, overlap = TRUE, ext = NULL)
temp2 <- merge(reproject_brick_201025, reproject_brick_20180517, tolerance
= 0.5, overlap = TRUE, ext = NULL)
temp3 <- merge(reproject_brick_20180524, reproject_brick_20190128, reproje
ct_brick_20190128_2, tolerance = 0.5, overlap = TRUE, extent = NULL)
temp_x <- merge(temp1, temp2, tolerance = 0.5, overlap = TRUE, ext = NULL)
plotRGB(reproject_brick_201023, r = 6, g = 5, b = 4, stretch = "lin", axes
= FALSE, main = "Vegetation Analysis", box = FALSE, add = TRUE) # Vegetati
on Analysis from bricks (NE UK)


# Adding Rede catchment layer ####
Rede_catchment <- read_sf("DATA/GIS_data/Rede_catchment/23008.shp")
Rede_catchment
plot(Rede_catchment$geometry, border = "green", add = TRUE)


#### Setting Communities dataset and convert to spatial ####

NVC_Communities <- read.csv("DATA/GIS_data/Communities/NVC_Communities.csv
") # read U1 community CSV file
NVC_Communities <- NVC_Communities[,-4] # removes open circle column
NVC_Communities <- NVC_Communities[,-6] # removes empty column at the end
colnames(NVC_Communities) <- c("Grid", "NVC", "Comm_Name", "Easting", "Nor
thing")

head(NVC_Communities)
## Subset data according to the communities of biodiversity importance und
er the UK BAP ####

# CG10 ####
CG10_community <- subset(NVC_Communities, NVC == "CG10") # CG10
CG10_spatial <- st_as_sf(CG10_community, coords = c("Easting", "Northing")
) # set coordinates (Eastings and Northings) and convert dataframe to simp
le feature
CG10_spatial_crs <- st_set_crs(CG10_spatial, 27700) # set projection to GB
OS
CG10_spatial_crs
CG10_england_wales <- st_intersection(CG10_spatial_crs, england_wales)
plot(st_geometry(CG10_england_wales), pch = 16, col = "blue", add = TRUE)

# CG11 ####
CG11_community <- subset(NVC_Communities, NVC == "CG11") # CG11
CG11_spatial <- st_as_sf(CG11_community, coords = c("Easting", "Northing")
) # set coordinates (Eastings and Northings) and convert dataframe to simp
le feature
CG11_spatial_crs <- st_set_crs(CG11_spatial, 27700) # set projection to GB
OS
CG11_spatial_crs
CG11_england_wales <- st_intersection(CG11_spatial_crs, england_wales)
plot(st_geometry(CG11_england_wales), pch = 16, col = "red", add = TRUE)
```

```
# CG12 ####
CG12_community <- subset(NVC_Communities, NVC == "CG12") # CG12
CG12_spatial <- st_as_sf(CG12_community, coords = c("Easting", "Northing")
) # set coordinates (Eastings and Northings) and convert dataframe to simp
le feature
CG12_spatial_crs <- st_set_crs(CG12_spatial, 27700) # set projection to GB
OS
CG12_spatial_crs
CG12_england_wales <- st_intersection(CG12_spatial_crs, england_wales)

# CG13 ####
CG13_community <- subset(NVC_Communities, NVC == "CG13") # CG13
CG13_spatial <- st_as_sf(CG13_community, coords = c("Easting", "Northing")
) # set coordinates (Eastings and Northings) and convert dataframe to simp
le feature
CG13_spatial_crs <- st_set_crs(CG13_spatial, 27700) # set projection to GB
OS
CG13_spatial_crs
CG13_england_wales <- st_intersection(CG13_spatial_crs, england_wales)

# CG14 ####
CG14_community <- subset(NVC_Communities, NVC == "CG14") # CG14
CG14_spatial <- st_as_sf(CG14_community, coords = c("Easting", "Northing")
) # set coordinates (Eastings and Northings) and convert dataframe to simp
le feature
CG14_spatial_crs <- st_set_crs(CG14_spatial, 27700) # set projection to GB
OS
CG14_spatial_crs
CG14_england_wales <- st_intersection(CG14_spatial_crs, england_wales)

#  CG9 ####
CG9_community <- subset(NVC_Communities, NVC == "CG9") # CG9
CG9_spatial <- st_as_sf(CG9_community, coords = c("Easting", "Northing"))
# set coordinates (Eastings and Northings) and convert dataframe to simple
feature
CG9_spatial_crs <- st_set_crs(CG9_spatial, 27700) # set projection to GB O
S
CG9_spatial_crs
CG9_england_wales <- st_intersection(CG9_spatial_crs, england_wales)

soil_1k <- raster("DATA/GIS_data/Rasters/soil_1k.tif")
DEM_1k <- raster("DATA/GIS_data/Rasters/DEM_1k.tif") # used as template fo
r resampling
aspect_1k <- raster("DATA/GIS_data/Rasters/aspect_1k.tif")
slope_1k <- raster("DATA/GIS_data/Rasters/slope_1k.tif")
LCM2015_1k <- raster("DATA/GIS_data/Rasters/LCM2015_1k.tif")
rainfall_1k <- raster("DATA/GIS_data/Rasters/rainfall.tif")
tempmin_summer_MET <- raster("DATA/GIS_data/Rasters/tempmin_summer_MET.tif
")
tempmin_winter_MET <- raster("DATA/GIS_data/Rasters/tempmin_winter_MET.tif
")
tempmax_summer_MET <- raster("DATA/GIS_data/Rasters/tempmax_summer_MET.tif
")
tempmax_winter_MET <- raster("DATA/GIS_data/Rasters/tempmax_winter_MET.tif
")
```

```
LANDSAT_NIR_1k <- raster("DATA/GIS_data/Rasters/LANDSAT_NIR.tif")
LANDSAT_Red_1k <- raster("DATA/GIS_data/Rasters/LANDSAT_Red.tif")
LANDSAT_Blue_1k <- raster("DATA/GIS_data/Rasters/LANDSAT_Blue.tif")

LANDSAT_EVI <- EVI(a=LANDSAT_NIR_1k, b=LANDSAT_Red_1k, c=LANDSAT_Blue_1k,
Pixel.Depth = 1)
writeRaster(LANDSAT_EVI, "DATA/GIS_data/Rasters/EVI_1k.tif")

CG10 <- st_read("DATA/GIS_data/Communities/Shapefiles/cropped/CG10/CG10.sh
p")
CG10 <- st_transform(CG10, 27700)
CG11 <- st_read("DATA/GIS_data/Communities/Shapefiles/cropped/CG11/CG11.sh
p")
CG11 <- st_transform(CG11, 27700)
CG12 <- st_read("DATA/GIS_data/Communities/Shapefiles/cropped/CG12/CG12.sh
p")
CG12 <- st_transform(CG12, 27700)
CG13 <- st_read("DATA/GIS_data/Communities/Shapefiles/cropped/CG13/CG13.sh
p")# no
CG13 <- st_transform(CG13, 27700)
CG14 <- st_read("DATA/GIS_data/Communities/Shapefiles/cropped/CG14/CG14.sh
p")
CG14 <- st_transform(CG14, 27700)
CG9 <- st_read("DATA/GIS_data/Communities/Shapefiles/cropped/CG9/CG9.shp")
CG9 <- st_transform(CG9, 27700)

####CG10_predict####
CG10_sp <- as(CG10, "Spatial")
CG10_sp$CG10 <- 1
#temp <- sdmData(pres ~ LCM2015_new + soil_new, train = CG10, predictors =
full_stack)
set.seed(100)
CG10_sdmData <- sdmData(CG10 ~  aspect_1k + DEM_1k + slope_1k + LCM2015_1k
+ layer
                        + soil_1k + rainfall + tempmin_summer_MET + tempmi
n_winter_MET + tempmax_summer_MET + tempmax_winter_MET,
                        train = CG10_sp, predictors = full_stack_1k, bg =
list(n=100, method = 'gRandom', remove = TRUE))
#temp_sdm <- sdm(pres ~ LCM2015_new + soil_new, data = , methods = c('glm'
, 'rf'), replication = NULL, n = 10)
#predict_temp <- predict(temp_sdm, full_stack, mean = T)

set.seed(000)
CG10_sdm <- sdm(CG10 ~ aspect_1k + DEM_1k + slope_1k + LCM2015_1k + layer
                + soil_1k + rainfall + tempmin_summer_MET + tempmin_winter
_MET + tempmax_summer_MET + tempmax_winter_MET,
                data = CG10_sdmData, methods = c('glm', 'svm', 'rf', 'entr
opy', 'maxlike'), replication = 'boot', n = 10)
#setting <- sdmSetting(pres~., methods = c('glm', 'gam', 'brt', 'svm', 'rf
'), replication = NULL, test.percent = 30, n = 10, modelSettings = list(br
t = list(n.trees = 500)))
predict_CG10_mean <- predict(CG10_sdm, full_stack_1k, overwrite = TRUE, me
an = T, filename = "DATA/GIS_data/Rasters/Predict/CG10_mean_predict.img")
predict_CG10 <- predict(CG10_sdm, full_stack_1k, filename = "DATA/GIS_data
/Rasters/Predict/CG10_predict.img")
```

```
#### CG11_predict ####

CG11_sp <- as(CG11, "Spatial")
CG11_sp$CG11 <- 1
#temp <- sdmData(pres ~ LCM2015_new + soil_new, train = CG11, predictors =
full_stack)
set.seed(101)
CG11_sdmData <- sdmData(CG11 ~  aspect_1k + DEM_1k + slope_1k + LCM2015_1k
+ layer
                        + soil_1k + rainfall + tempmin_summer_MET + tempmi
n_winter_MET + tempmax_summer_MET + tempmax_winter_MET,
                        train = CG11_sp, predictors = full_stack_1k, bg =
list(n=100, method = 'gRandom', remove = TRUE))
#temp_sdm <- sdm(pres ~ LCM2015_new + soil_new, data = , methods = c('glm'
, 'rf', 'entropy', 'maxlike'), replication = NULL, n = 10)
#predict_temp <- predict(temp_sdm, full_stack, mean = T)

set.seed(001)
CG11_sdm <- sdm(CG11 ~ aspect_1k + DEM_1k + slope_1k + LCM2015_1k + layer
                + soil_1k + rainfall + tempmin_summer_MET + tempmin_winter
_MET + tempmax_summer_MET + tempmax_winter_MET,
                data = CG11_sdmData, methods = c('glm', 'svm', 'rf', 'entr
opy', 'maxlike'), replication = 'boot', n = 10)
#setting <- sdmSetting(pres~., methods = c('glm', 'gam', 'brt', 'svm', 'rf
', 'entropy', 'maxlike'), replication = NULL, test.percent = 30, n = 10, m
odelSettings = list(brt = list(n.trees = 500)))
predict_CG11_mean <- predict(CG11_sdm, full_stack_1k, overwrite = TRUE, me
an = T, filename = "DATA/GIS_data/Rasters/Predict/CG11_mean_predict.img")
predict_CG11 <- predict(CG11_sdm, full_stack_1k, overwrite = TRUE, filenam
e = "DATA/GIS_data/Rasters/Predict/CG11_predict.img")

#### CG12 ####

#### CG12_predict ####
set.seed(102)
CG12_sp <- as(CG12, "Spatial")
CG12_sp$CG12 <- 1
#temp <- sdmData(pres ~ LCM2015_new + soil_new, train = CG12, predictors =
full_stack)
CG12_sdmData <- sdmData(CG12 ~  aspect_1k + DEM_1k + slope_1k + LCM2015_1k
+ layer
                        + soil_1k + rainfall + tempmin_summer_MET + tempmi
n_winter_MET + tempmax_summer_MET + tempmax_winter_MET,
                        train = CG12_sp, predictors = full_stack_1k, bg =
list(n=100, method = 'gRandom', remove = TRUE))
#temp_sdm <- sdm(pres ~ LCM2015_new + soil_new, data = , methods = c('glm'
, 'rf', 'entropy', 'maxlike'), replication = NULL, n = 10)
#predict_temp <- predict(temp_sdm, full_stack, mean = T)

set.seed(002)
CG12_sdm <- sdm(CG12 ~ aspect_1k + DEM_1k + slope_1k + LCM2015_1k + layer
                + soil_1k + rainfall + tempmin_summer_MET + tempmin_winter
_MET + tempmax_summer_MET + tempmax_winter_MET,
                data = CG12_sdmData, methods = c('glm', 'svm', 'rf', 'entr
opy', 'maxlike'), replication = 'boot', n = 10)
#setting <- sdmSetting(pres~., methods = c('glm', 'gam', 'brt', 'svm', 'rf
```

```
', 'entropy', 'maxlike'), replication = NULL, test.percent = 30, n = 10, m
odelSettings = list(brt = list(n.trees = 500)))
predict_CG12_mean <- predict(CG12_sdm, full_stack_1k, overwrite = TRUE, me
an = T, filename = "DATA/GIS_data/Rasters/Predict/CG12_mean_predict.img")
predict_CG12 <- predict(CG12_sdm, full_stack_1k, filename = "DATA/GIS_data
/Rasters/Predict/CG12_predict.img")


#### CG14_predict ####

CG14_sp <- as(CG14, "Spatial")
CG14_sp$CG14 <- 1
#temp <- sdmData(pres ~ LCM2015_new + soil_new, train = CG14, predictors =
full_stack)
set.seed(103)
CG14_sdmData <- sdmData(CG14 ~  aspect_1k + DEM_1k + slope_1k + LCM2015_1k
+ layer
                        + soil_1k + rainfall + tempmin_summer_MET + tempmi
n_winter_MET + tempmax_summer_MET + tempmax_winter_MET,
                        train = CG14_sp, predictors = full_stack_1k, bg =
list(n=100, method = 'gRandom', remove = TRUE))
#temp_sdm <- sdm(pres ~ LCM2015_new + soil_new, data = , methods = c('glm'
, 'rf', 'entropy', 'maxlike'), replication = NULL, n = 10)
#predict_temp <- predict(temp_sdm, full_stack, mean = T)

set.seed(004)
CG14_sdm <- sdm(CG14 ~ aspect_1k + DEM_1k + slope_1k + LCM2015_1k + layer
                + soil_1k + rainfall + tempmin_summer_MET + tempmin_winter
_MET + tempmax_summer_MET + tempmax_winter_MET,
                data = CG14_sdmData, methods = c('glm', 'svm', 'rf', 'entr
opy', 'maxlike'), replication = 'boot', n = 10)
#setting <- sdmSetting(pres~., methods = c('glm', 'gam', 'brt', 'svm', 'rf
', 'entropy', 'maxlike'), replication = NULL, test.percent = 30, n = 10, m
odelSettings = list(brt = list(n.trees = 500)))
predict_CG14_mean <- predict(CG14_sdm, full_stack_1k, overwrite = TRUE, me
an = T, filename = "DATA/GIS_data/Rasters/Predict/CG14_mean_predict.img")
predict_CG14 <- predict(CG14_sdm, full_stack_1k, filename = "DATA/GIS_data
/Rasters/Predict/CG14_predict.img")


#### CG9_predict ####

CG9_sp <- as(CG9, "Spatial")
CG9_sp$CG9 <- 1
#temp <- sdmData(pres ~ LCM2015_new + soil_new, train = CG9, predictors =
full_stack)
set.seed(104)
CG9_sdmData <- sdmData(CG9 ~  aspect_1k + DEM_1k + slope_1k + LCM2015_1k +
layer
                        + soil_1k + rainfall + tempmin_summer_MET + tempmi
n_winter_MET + tempmax_summer_MET + tempmax_winter_MET,
                        train = CG9_sp, predictors = full_stack_1k, bg = l
ist(n=100, method = 'gRandom', remove = TRUE))
#temp_sdm <- sdm(pres ~ LCM2015_new + soil_new, data = , methods = c('glm'
, 'rf', 'entropy', 'maxlike'), replication = NULL, n = 10)
#predict_temp <- predict(temp_sdm, full_stack, mean = T)
```

```
set.seed(005)
CG9_sdm <- sdm(CG9 ~ aspect_1k + DEM_1k + slope_1k + LCM2015_1k + layer
               + soil_1k + rainfall + tempmin_summer_MET + tempmin_winter
_MET + tempmax_summer_MET + tempmax_winter_MET,
               data = CG9_sdmData, methods = c('glm', 'svm', 'rf', 'entro
py', 'maxlike'), replication = 'boot', n = 10)
#setting <- sdmSetting(pres~., methods = c('glm', 'gam', 'brt', 'svm', 'rf
', 'entropy', 'maxlike'), replication = NULL, test.percent = 30, n = 10, m
odelSettings = list(brt = list(n.trees = 500)))
predict_CG9_mean <- predict(CG9_sdm, full_stack_1k, overwrite = TRUE, mean
= T, filename = "DATA/GIS_data/Rasters/Predict/CG9_mean_predict.img")
predict_CG9 <- predict(CG9_sdm, full_stack_1k, filename = "DATA/GIS_data/R
asters/Predict/CG9_predict.img")

#########
#### Model Evaluation ####
# To get AUC, TSS and Kappa

#Using CG10 as test data
CG10_eval <- getEvaluation(CG10_sdm, stat = c('TSS', 'Kappa', 'AUC'), opt
= 1)
#p1_CG10 <- ensemble(CG10_sdm, newdata = full_stack_1k, filename = "ensCG1
0.grd", overwrite = TRUE, setting = list(method = 'weighted', stat = 'Kapp
a'))
#CG10#_niche <- niche(x = full_stack_1k, h = #p1_CG10, c('soil_1k', 'LCM20
15_1k'))
CG10_eval_mean <- colMeans(CG10_eval)

CG11_eval <- getEvaluation(CG11_sdm, stat = c('TSS', 'Kappa', 'AUC'), opt
= 1)
#p1_CG11 <- ensemble(CG11_sdm, newdata = full_stack_1k, filename = "ensCG1
1.grd", overwrite = TRUE, setting = list(method = 'weighted', stat = 'Kapp
a'))
#CG11#_niche <- niche(x = full_stack_1k, h = #p1_CG11, c('soil_1k', 'LCM20
15_1k'))
CG11_eval_mean <- colMeans(CG11_eval)

CG12_eval <- getEvaluation(CG12_sdm, stat = c('TSS', 'Kappa', 'AUC'), opt
= 1)
#p1_CG12 <- ensemble(CG12_sdm, newdata = full_stack_1k, filename = "ensCG1
2.grd", overwrite = TRUE, setting = list(method = 'weighted', stat = 'Kapp
a'))
#CG12#_niche <- niche(x = full_stack_1k, h = #p1_CG12, c('soil_1k', 'LCM20
15_1k'))
CG12_eval_mean <- colMeans(CG12_eval)

#CG13_eval <- getEvaluation(CG13_sdm, stat = c('TSS', 'Kappa', 'AUC'), opt
= 1)
##p1_CG13 <- ensemble(CG13_sdm, newdata = full_stack_1k, filename = "ensCG
13.grd", overwrite = TRUE, setting = list(method = 'weighted', stat = 'Kap
pa'))
#CG13#_niche <- niche(x = full_stack_1k, h = #p1_CG13, c('soil_1k', 'LCM20
15_1k'))

CG14_eval <- getEvaluation(CG14_sdm, stat = c('TSS', 'Kappa', 'AUC'), opt
```

```
= 1)
#p1_CG14 <- ensemble(CG14_sdm, newdata = full_stack_1k, filename = "ensCG1
4.grd", overwrite = TRUE, setting = list(method = 'weighted', stat = 'Kapp
a'))
#CG14#_niche <- niche(x = full_stack_1k, h = #p1_CG14, c('soil_1k', 'LCM20
15_1k'))
CG14_eval_mean <- colMeans(CG14_eval)

CG9_eval <- getEvaluation(CG9_sdm, stat = c('TSS', 'Kappa', 'AUC'), opt =
1)
#p1_CG9 <- ensemble(CG9_sdm, newdata = full_stack_1k, filename = "ensCG9.g
rd", overwrite = TRUE, setting = list(method = 'weighted', stat = 'Kappa')
)
#CG9#_niche <- niche(x = full_stack_1k, h = #p1_CG9, c('soil_1k', 'LCM2015
_1k'))
CG9_eval_mean <- colMeans(CG9_eval)

#### Evaluation of models_mean ####
# To get barplot of best statistical model used in SDM

grouped <- read.csv("DATA/GIS_data/SDM/var_imp/Model_evaluations/for_plot/
grouped.csv")
plot_grouped <- ggplot(grouped_se, aes(fill = Stat_type, y = Per, x = Stat
_type )) +
  geom_bar(position = "dodge", stat = "identity") + geom_errorbar(aes(ymin
= Per-SE_100, ymax = Per+SE_100),

width = 0.2, position = position_dodge(0.9))
plot_grouped  + scale_fill_grey(start = 0.1, end = 0.8, aesthetics = "fill
") +scale_y_continuous(limits = c(0, 100), expand = c(0, 0)) +
  facet_wrap(~ Condition) +
  theme_bw() + theme(
   plot.background = element_blank()
  ,panel.grid.major = element_blank()
  ,panel.grid.minor = element_blank()
  #,axis.text.x = element_blank()
  #,axis.ticks.x = element_blank()
 #,panel.border = element_blank()
 ,axis.line = element_blank()
 #,axis.text.x = element_blank()
 #,axis.text.y = element_blank()
 #,axis.ticks.x = element_blank()
 #,axis.ticks.y = element_blank()
 ,axis.title = element_blank()
 , legend.position = "none")

## function for std.error
s.err <- function(x) sqrt(var(x)/length(x))

GLM_results <- read.csv("DATA/GIS_data/SDM/var_imp/Model_evaluations/for_p
lot/GLM.csv")
SVM_results <- read.csv("DATA/GIS_data/SDM/var_imp/Model_evaluations/for_p
lot/SVM.csv")
RF_results <- read.csv("DATA/GIS_data/SDM/var_imp/Model_evaluations/for_pl
ot/RF.csv")
MAXENT_results <-read.csv("DATA/GIS_data/SDM/var_imp/Model_evaluations/for
```

```
_plot/MAXENT.csv")
MAXLIKE_results <- read.csv("DATA/GIS_data/SDM/var_imp/Model_evaluations/f
or_plot/MAXLIKE.csv")
GLM_results <- GLM_results[-54,]
SVM_results <- SVM_results[-54,]
RF_results <- RF_results[-54,]
MAXENT_results <- MAXENT_results[-54,]
MAXLIKE_results <- MAXLIKE_results[-54,]


stand.err <- NULL
stand.error <- data.frame(rbind(s.err(GLM_results[,2]), s.err(SVM_results[
,2]), s.err(RF_results[,2]), s.err(MAXENT_results[,2]), s.err(MAXLIKE_resu
lts[,2]),
                                s.err(GLM_results[,3]), s.err(SVM_results[
,3]), s.err(RF_results[,3]), s.err(MAXENT_results[,3]), s.err(MAXLIKE_resu
lts[,3]),
                                s.err(GLM_results[,4]), s.err(SVM_results[
,4]), s.err(RF_results[,4]), s.err(MAXENT_results[,4]), s.err(MAXLIKE_resu
lts[,4])))
stand_error_100 <- stand.error*100
grouped$SE100 <- stand_error_100
colnames(grouped) <- c("Stat_type", "Condition", "Value", "Per", "Stderr",
"S.E.100")

grouped_se <- read.csv("DATA/GIS_data/SDM/var_imp/Model_evaluations/for_pl
ot/group_w_se.csv")

Stats_types <- read.csv("DATA/GIS_data/SDM/var_imp/Model_evaluations/for_p
lot/Stats_types.csv")
colnames(Stats_types) <- c("Community", "Type", "AUC", "AUC_val", "COR", "
COR_val", "TSS", "TSS_val")

### AUC boxplots

AUC_plots <- ggplot(Stats_types, aes(x = Type, y = AUC_val)) + geom_boxplo
t()
AUC_plots + theme_bw() + theme(
    plot.background = element_blank()
    ,panel.grid.major = element_blank()
    ,panel.grid.minor = element_blank()
    #,axis.text.x = element_blank()
    #,axis.ticks.x = element_blank()
    ,panel.border = element_blank()
    ,axis.line = element_blank()
    #,axis.text.x = element_blank()
    #,axis.text.y = element_blank()
    #,axis.ticks.x = element_blank()
    #,axis.ticks.y = element_blank()
    ,axis.title = element_blank()
    ,axis.line.x = element_line("black")
    ,axis.line.y = element_line("black")
)
### COR boxplots
COR_plots <- ggplot(Stats_types, aes(x = Type, y = COR_val)) + geom_boxplo
t()
```

```
COR_plots + theme_bw() + theme(
  plot.background = element_blank()
  ,panel.grid.major = element_blank()
  ,panel.grid.minor = element_blank()
  #,axis.text.x = element_blank()
  #,axis.ticks.x = element_blank()
  ,panel.border = element_blank()
  ,axis.line = element_blank()
  #,axis.text.x = element_blank()
  #,axis.text.y = element_blank()
  #,axis.ticks.x = element_blank()
  #,axis.ticks.y = element_blank()
  ,axis.title = element_blank()
  ,axis.line.x = element_line("black")
  ,axis.line.y = element_line("black")
)

### TSS boxplots
TSS_plots <- ggplot(Stats_types, aes(x = Type, y = TSS_val)) + geom_boxplo
t()
TSS_plots + theme_bw() + theme(
  plot.background = element_blank()
  ,panel.grid.major = element_blank()
  ,panel.grid.minor = element_blank()
  #,axis.text.x = element_blank()
  #,axis.ticks.x = element_blank()
  ,panel.border = element_blank()
  ,axis.line = element_blank()
  #,axis.text.x = element_blank()
  #,axis.text.y = element_blank()
  #,axis.ticks.x = element_blank()
  #,axis.ticks.y = element_blank()
  ,axis.title = element_blank()
  ,axis.line.x = element_line("black")
  ,axis.line.y = element_line("black")
)

# Variable importance

CG10_var <- getVarImp(CG10_sdm, id = 1, wtest = 'training')
CG11_var <- getVarImp(CG11_sdm, id = 1, wtest = 'training')
CG12_var <- getVarImp(CG12_sdm, id = 1, wtest = 'training')
#CG13_var <- getVarImp(CG13_sdm, id = 1, wtest = 'training')
CG14_var <- getVarImp(CG14_sdm, id = 1, wtest = 'training')
CG9_var <- getVarImp(CG9_sdm, id = 1, wtest = 'training')
var_imp <- rbind(CG10_var, CG11_var, CG12_var, CG14_var, CG9_var, H10_var
                 H12_var, H13_var, H14_var, H15_var, H16_var, H18_var,
                 H19_var, H21_var, H22_var, H8_var, H9_var, M1_var,
                 M11_var, M15_var, M16_var, M17_var, M18_var, M19_var,
                 M2_var, M20_var, M21_var, M25_var, M26_var, M27_var,
                 M29_var, M3_var, M37_var, M38_var, M4_var, M5_var,
                 M8_var, M9_var, MG10_var, MG5_var, U10_var, U13_var,
                 U15_var, U16_var, U17_var, U2_var, U21_var, U4_var,
                 U5_var, U6_var, U7_var, W19_var, W9_var)

#### Plot predicted means of sdm ####
pdf("DATA/GIS_data/PDFs/predictions_EVI.pdf")
```

```
#plot(predict_CG10, main = "CG10")
plot(predict_CG10_mean, axes = FALSE, box = FALSE)#, "DATA/GIS_data/PDFs/C
G10/CG10_predict_mean.pdf")
#plot(predict_CG11, main = "CG11", axes = FALSE, box = FALSE)#, "DATA/GIS_
data/PDFs/CG11/CG11_predict.pdf")
plot(predict_CG11_mean, axes = FALSE, box = FALSE)#, "DATA/GIS_data/PDFs/C
G11/CG11_predict_mean.pdf")
#plot(predict_CG12, main = "CG12", axes = FALSE, box = FALSE)#, "DATA/GIS_
data/PDFs/CG12/CG12_predict.pdf")
plot(predict_CG12_mean, axes = FALSE, box = FALSE)#, "DATA/GIS_data/PDFs/C
G12/CG12_predict_mean.pdf")
#plot(predict_CG14, main = "CG14", axes = FALSE, box = FALSE)#, "DATA/GIS_
data/PDFs/CG14/CG14_predict.pdf")
plot(predict_CG14_mean, axes = FALSE, box = FALSE)#, "DATA/GIS_data/PDFs/C
G14/CG14_predict_mean.pdf")
#plot(predict_CG9, main = "CG9", axes = FALSE, box = FALSE)#, "DATA/GIS_da
ta/PDFs/CG9/CG9_predict.pdf")
plot(predict_CG9_mean, axes = FALSE, box = FALSE)#, "DATA/GIS_data/PDFs/CG
9/CG9_predict_mean.pdf")

#### NBN Atlas ####
# Caching data
nbn_config(cache_directory = "DATA/GIS_data/NBN_cache")
setHook(packageEvent("NBN4R", "attach"), function(...) nbn_config(cache_di
rectory = file.path("~","NBN_cache_data", "NBN_cache")))

#### C. vulgaris ####
calluna <- search_fulltext("Calluna vulgaris")
call_occ <- occurrences(taxon = "Calluna vulgaris", download_reason_id = 4
)
call_df <- data.frame(call_occ$data)
call_df_main <- data.frame(cbind(call_df$scientificName), call_df$rank, ca
ll_df$occurrenceStatus, call_df$locality, call_df$OSGR, call_df$latitudeWG
S84,
                           call_df$longitudeWGS84, call_df$OSGR10km, call_
df$OSGR1km, call_df$individualCount, call_df$abundance, call_df$OSGR, call
_df$decimalLatLongCalculatedFromGridReference, call_df$stateProvince)
call_df_engwal <- subset(call_df_main, call_df.stateProvince!="Scotland")
call_df_engwal <- subset(call_df_engwal, call_df.stateProvince!="Isle of M
an")
call_df_engwal <- subset(call_df_engwal, call_df.stateProvince!="Northern
Ireland")
call_df_engwal <- call_df_engwal[complete.cases(call_df_engwal[, 6:7]),] #
remove rows from lat/long with NAs

head(call_occ$data)

call_engwal <- st_as_sf(call_df_engwal, coords = c("call_df.longitudeWGS84
", "call_df.latitudeWGS84"), crs = crs_latlong)
call_engwal <- st_transform(call_engwal, 27700)
call_engwal_crop <- st_crop(call_engwal, england_wales)
st_write(call_engwal_crop, "DATA/GIS_data/NBN_atlas/call_vulg.shp")

#### U2 species prediction ####
U2_rede_mask
Rede <- as(Rede_roy, 'Spatial')
```

```
U2_glm <- as.data.frame(U2_rede_mask$sp_1.m_glm.re_boot, xy = TRUE)


rede_gg <- ggplot(Rede_roy) + geom_sf() + geom_raster(data = U2_glm, aes(x
=x, y=y, alpha = sp_1.m_glm.re_boot, fill = sp_1.m_glm.re_boot))
descflex_gg <- ggplot(descflex, aes(x=X, y=Y)) + geom_point(aes(x=X, y=Y))
callvulg_gg <- ggplot(callvulg) + geom_point(aes(x=X, y=Y))

 rede_gg + geom_jitter(data = descflex, aes(x=X, y=Y), position = position
_jitter(width = 0, height = 0.5)) + theme_bw() +
   theme(
    plot.background = element_blank()
    ,panel.grid.major.x = element_blank()
    ,panel.grid.minor.x = element_blank()
    ,panel.grid.major.y = element_blank()
    ,panel.grid.minor.y = element_blank()
    ,panel.background = element_blank()
    ,panel.border = element_blank()
    ,axis.line = element_blank()
    ,axis.text.x = element_blank()
    ,axis.text.y = element_blank()
    ,axis.ticks.x = element_blank()
    ,axis.ticks.y = element_blank()
    ,axis.title = element_blank()
    )

 tempgeom_jitter(data = callvulg, aes(x=X, y=Y), size = 1, position=positi
on_jitter(width = 0, height = 0.5))
```

“Because I was standing in the tunnel. And I was really there.

And that was enough to make me feel infinite.”

- Stephen Chbosky, *The Perks of Being a Wallflower*